### (19) World Intellectual Property Organization International Bureau



### 

### (43) International Publication Date 28 June 2001 (28.06.2001)

### **PCT**

# (10) International Publication Number WO 01/45748 A1

(51) International Patent Classification7: C12Q 1/70 A61K 48/00,

(21) International Application Number: PCT/US00/34724

(22) International Filing Date:

21 December 2000 (21.12.2000)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data: 60/171,542

22 December 1999 (22.12.1999) US

(71) Applicant (for all designated States except US): MERCK & CO., INC. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): SHIVER, John, W. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). PERRY, Helen, C. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). CASIMIRO, Danilo, R. [PH/US]; 126 East Lincoln Avenue, Rahway,

NJ 07065-0907 (US). FU, Tong-Ming [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US).

- (74) Common Representative: MERCK & CO., INC.; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

#### Published

With international search report.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: POLYNUCLEOTIDE VACCINES EXPRESSING CODON OPTIMIZED HIV-1 POL AND MODIFIED HIV-1 POL

(57) Abstract: Pharmaceutical compositions which comprise HIV Pol DNA vaccines are disclosed, along with the production and use of these DNA vaccines. The pol-based DNA vaccines of the invention are administered directly introduced into living vertebrate tissue, preferably humans, and preferably express inactivated versions of the HIV Pol protein devoid of protease, reverse transcriptase activity, RNase H activity and integrase activity, inducing a cellular immune response which specifically recognizes human immunodeficiency virus-1 (HIV-1). The DNA molecules which comprise the open reading frame of these DNA vaccines are synthetic DNA molecules encoding codon optimized HIV-1 Pol and codon optimized inactive derivatives of optimized HIV-1 Pol, including DNA molecules which encode inactive Pol proteins which comprise an amino terminal leader peptide.

TITLE OF THE INVENTION
POLYNUCLEOTIDE VACCINES EXPRESSING CODON OPTIMIZED HIV-1
POL AND MODIFIED HIV-1 POL

### CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit, under 35 U.S.C. §119(e), of U.S. provisional application 60/171,542, filed December 22, 1999.

10

5

## STATEMENT REGARDING FEDERALLY-SPONSORED R&D Not Applicable

15 REFERENCE TO MICROFICHE APPENDIX
Not Applicable

### FIELD OF THE INVENTION

The present invention relates to HIV Pol polynucleotide pharmaceutical 20 products, as well as the production and use thereof which, when directly introduced into living vertebrate tissue, preferably a mammalian host such as a human or a non-human mammal of commercial or domestic veterinary importance, express the HIV Pol protein or biologically relevant portions thereof within the animal, inducing a cellular immune response which specifically recognizes human immunodeficiency 25 virus-1 (HIV-1). The polynucleotides of the present invention are synthetic DNA molecules encoding codon optimized HIV-1 Pol and derivatives of optimized HIV-1 Pol, including constructs wherein protease, reverse transcriptase, RNAse H and integrase activity of HIV-1 Pol is inactivated. The polynucleotide vaccines of the present invention should offer a prophylactic advantage to previously uninfected 30 individuals and/or provide a therapeutic effect by reducing viral load levels within an infected individual, thus prolonging the asymptomatic phase of HIV-1 infection.

### BACKGROUND OF THE INVENTION

10

15

20

25

30

Human Immunodeficiency Virus-1 (HIV-1) is the etiological agent of acquired human immune deficiency syndrome (AIDS) and related disorders. HIV-1 is an RNA virus of the Retroviridae family and exhibits the 5'LTR-gag-pol-env-LTR 3'organization of all retroviruses. The integrated form of HIV-1, known as the provirus, is approximately 9.8 Kb in length. Each end of the viral genome contains flanking sequences known as long terminal repeats (LTRs). The HIV genes encode at least nine proteins and are divided into three classes; the major structural proteins (Gag, Pol, and Env), the regulatory proteins (Tat and Rev); and the accessory proteins (Vpu, Vpr, Vif and Nef).

The gag gene encodes a 55-kilodalton (kDa) precursor protein (p55) which is expressed from the unspliced viral mRNA and is proteolytically processed by the HIV protease, a product of the pol gene. The mature p55 protein products are p17 (matrix), p24 (capsid), p9 (nucleocapsid) and p6.

The *pol* gene encodes proteins necessary for virus replication; a reverse transcriptase, a protease, integrase and RNAse H. These viral proteins are expressed as a Gag-Pol fusion protein, a 160 kDa precursor protein which is generated via a ribosomal frame shifting. The viral encoded protease proteolytically cleaves the Pol polypeptide away from the Gag-Pol fusion and further cleaves the Pol polypeptide to the mature proteins which provide protease (Pro, P10), reverse transcriptase (RT, P50), integrase (IN, p31) and RNAse H (RNAse, p15) activities.

The *nef* gene encodes an early accessory HIV protein (Nef) which has been shown to possess several activities such as down regulating CD4 expression, disturbing T-cell activation and stimulating HIV infectivity.

The *env* gene encodes the viral envelope glycoprotein that is translated as a 160-kilodalton (kDa) precursor (gp160) and then cleaved by a cellular protease to yield the external 120-kDa envelope glycoprotein (gp120) and the transmembrane 41-kDa envelope glycoprotein (gp41). Gp120 and gp41 remain associated and are displayed on the viral particles and the surface of HIV-infected cells.

The *tat* gene encodes a long form and a short form of the Tat protein, a RNA binding protein which is a transcriptional transactivator essential for HIV-1 replication.

The rev gene encodes the 13 kDa Rev protein, a RNA binding protein. The Rev protein binds to a region of the viral RNA termed the Rev response element

(RRE). The Rev protein is promotes transfer of unspliced viral RNA from the nucleus to the cytoplasm. The Rev protein is required for HIV late gene expression and in turn, HIV replication.

5

10

15

20

25

30

Gp120 binds to the CD4/chemokine receptor present on the surface of helper T-lymphocytes, macrophages and other target cells in addition to other co-receptor molecules. X4 (macrophage tropic) virus show tropism for CD4/CXCR4 complexes while a R5 (T-cell line tropic) virus interacts with a CD4/CCR5 receptor complex. After gp120 binds to CD4, gp41 mediates the fusion event responsible for virus entry. The virus fuses with and enters the target cell, followed by reverse transcription of its single stranded RNA genome into the double-stranded DNA via a RNA dependent DNA polymerase. The viral DNA, known as provirus, enters the cell nucleus, where the viral DNA directs the production of new viral RNA within the nucleus, expression of early and late HIV viral proteins, and subsequently the production and cellular release of new virus particles. Recent advances in the ability to detect viral load within the host shows that the primary infection results in an extremely high generation and tissue distribution of the virus, followed by a steady state level of virus (albeit through a continual viral production and turnover during this phase), leading ultimately to another burst of virus load which leads to the onset of clinical AIDS. Productively infected cells have a half life of several days, whereas chronically or latently infected cells have a 3-week half life, followed by non-productively infected cells which have a long half life (over 100 days) but do not significantly contribute to day to day viral loads seen throughout the course of disease.

Destruction of CD4 helper T lymphocytes, which are critical to immune defense, is a major cause of the progressive immune dysfunction that is the hallmark of HIV infection. The loss of CD4 T-cells seriously impairs the body's ability to fight most invaders, but it has a particularly severe impact on the defenses against viruses, fungi, parasites and certain bacteria, including mycobacteria.

Effective treatment regimens for HIV-1 infected individuals have become available recently. However, these drugs will not have a significant impact on the disease in many parts of the world and they will have a minimal impact in halting the spread of infection within the human population. As is true of many other infectious diseases, a significant epidemiologic impact on the spread of HIV-1 infection will only occur subsequent to the development and introduction of an effective vaccine. There are a number of factors that have contributed to the lack of successful vaccine

5

10

15

20

30

development to date. As noted above, it is now apparent that in a chronically infected person there exists constant virus production in spite of the presence of anti-HIV-1 humoral and cellular immune responses and destruction of virally infected cells. As in the case of other infectious diseases, the outcome of disease is the result of a balance between the kinetics and the magnitude of the immune response and the pathogen replicative rate and accessibility to the immune response. Pre-existing immunity may be more successful with an acute infection than an evolving immune response can be with an established infection. A second factor is the considerable genetic variability of the virus. Although anti-HTV-1 antibodies exist that can neutralize HIV-1 infectivity in cell culture, these antibodies are generally virus isolate-specific in their activity. It has proven impossible to define serological groupings of HIV-1 using traditional methods. Rather, the virus seems to define a serological "continuum" so that individual neutralizing antibody responses, at best, are effective against only a handful of viral variants. Given this latter observation, it would be useful to identify immunogens and related delivery technologies that are likely to elicit anti-HIV-1 cellular immune responses. It is known that in order to generate CTL responses antigen must be synthesized within or introduced into cells, subsequently processed into small peptides by the proteasome complex, and translocated into the endoplasmic reticulum/Golgi complex secretory pathway for eventual association with major histocompatibility complex (MHC) class I proteins. CD8<sup>+</sup> T lymphocytes recognize antigen in association with class I MHC via the T cell receptor (TCR) and the CD8 cell surface protein. Activation of naive CD8<sup>+</sup> T cells into activated effector or memory cells generally requires both TCR engagement of antigen as described above as well as engagement of costimulatory proteins. Optimal induction of CTL responses usually requires "help" in the form of cytokines from CD4<sup>+</sup> T lymphocytes which recognize antigen associated with MHC class II molecules via TCR and CD4 engagement.

Larder, et al., (1987, *Nature* 327: 716-717) and Larder, et al., (1989, *Proc. Natl. Acad. Sci.* 86: 4803-4807) disclose site specific mutagenesis of HIV-1 RT and the effect such changes have on *in vitro* activity and infectivity related to interaction with known inhibitors of RT.

Davies, et al. (1991, *Science* 252:, 88-95) disclose the crystal structure of the RNase H domain of HIV-1 Pol.

Schatz, et al. (1989, *FEBS Lett.* 257: 311-314) disclose that mutations Glu478Gln and His539Phe in a complete HIV-1 RT/RNase H DNA fragment results in defective RNase activity without effecting RT activity.

Mizrahi, et al. (1990, *Nucl. Acids. Res.* 18: pp. 5359-5353) disclose additional mutations Asp443Asn and Asp498Asn in the RNase region of the *pol* gene which also results in defective RNase activity. The authors note that the Asp498Asn mutant was difficult to characterize due to instability of this mutant protein.

Leavitt, et al. (1993, *J. Biol. Chem.* 268: 2113-2119) disclose several mutations, including a Asp64Val mutation, which show differing effect on HIV-1 integrase (IN) activity.

Wiskerchen, et al. (1995, *J. Virol*. 69: 376-386) disclose singe and double mutants, including mutation of aspartic acid residues which effect HIV-1 IN and viral replication functions.

It would be of great import in the battle against AIDS to produce a prophylactic- and/or therapeutic-based HIV vaccine which generates a strong cellular immune response against an HIV infection. The present invention addresses and meets this needs by disclosing a class of DNA vaccines based on host delivery and expression of modified versions of the HIV-1 gene, pol.

### 20 SUMMARY OF THE INVENTION

5

10

25

The present invention relates to synthetic DNA molecules (also referred to herein as "polynucleotides") and associated DNA vaccines (also referred to herein as "polynucleotide vaccines") which elicit cellular immune and humoral responses upon administration to the host, including primates and especially humans, and also including a non-human mammal of commercial or domestic veterinary importance. An effect of the cellular immune-directed vaccines of the present invention should be the lower transmission rate to previously uninfected individuals and/or reduction in the levels of the viral loads within an infected individual, so as to prolong the asymptomatic phase of HIV-1 infection. In particular, the present invention relates to DNA vaccines which encode various forms of HIV-1 Pol, wherein administration, intracellular delivery and expression of the HIV-1 Pol gene of interest elicits a host CTL and Th response. The preferred synthetic DNA molecules of the present invention encode codon optimized versions of wild type HIV-1 Pol, codon optimized versions of HIV-1 Pol fusion proteins, and codon optimized versions of HIV-1 Pol

proteins and fusion protein, including but not limited to *pol* modifications involving residues within the catalytic regions responsible for RT, RNase and IN activity within the host cell.

A particular embodiment of the present invention relates to codon optimized wt-pol DNA constructs wherein DNA sequences encoding the protease (PR) activity are deleted, leaving codon optimized "wild type" sequences which encode RT (reverse transcriptase and RNase H activity) and IN integrase activity. The nucleotide sequence of a DNA molecule which encodes this protein is disclosed herein as SEQ ID NO:1 and the corresponding amino acid sequence of the expressed protein is disclosed herein as SEQ ID NO:2.

5

10

15

20

25

30

The present invention preferably relates to a HIV-1 DNA pol construct which is devoid of DNA sequences encoding any PR activity, as well as containing a mutation(s) which at least partially, and preferably substantially, abolishes RT, RNase and/or IN activity. One type of HIV-1 pol mutant may include but is not limited to a mutated DNA molecule comprising at least one nucleotide substitution which results in a point mutation which effectively alters an active site within the RT, RNase and/or IN regions of the expressed protein, resulting in at least substantially decreased enzymatic activity for the RT, RNase H and/or IN functions of HIV-1 Pol. In a preferred embodiment of this portion of the invention, a HIV-1 DNA pol construct contains a mutation or mutations within the Pol coding region which effectively abolishes RT, RNase H and IN activity. An especially preferable HIV-1 DNA pol construct in a DNA molecule which contains at least one point mutation which alters the active site of the RT, RNase H and IN domains of Pol, such that each activity is at least substantially abolished. Such a HIV-1 Pol mutant will most likely comprise at least one point mutation in or around each catalytic domain responsible for RT, RNase H and IN activity, respectfully. To this end, an especially preferred HIV-1 DNA pol construct is exemplified herein and contains nine codon substitution mutations which results in an inactivated Pol protein (IA Pol: SEQ ID NO:4, Figure 2A-C) which has no PR, RT, RNase or IN activity, wherein three such point mutations reside within each of the RT, RNase and IN catalytic domains. Any combination of the mutations disclosed herein may suitable and therefore may be utilized as an IA-Pol-based vaccine of the present invention. While addition and deletion mutations are contemplated and within the scope of the invention, the

preferred mutation is a point mutation resulting in a substitution of the wild type amino acid with an alternative amino acid residue.

Another aspect of the present invention is to generate HIV-1 Pol-based vaccine constructions which comprise a eukaryotic trafficking signal peptide such as the leader peptide from human tPA. To this end, the present invention relates to a DNA molecule which encodes a codon optimized wt-pol DNA construct wherein the protease (PR) activity is deleted and a human tPA leader sequence is fused to the 5' end of the coding region. A DNA molecule which encodes this protein is disclosed herein as SEQ ID NO:5, the open reading frame disclosed herein as SEQ ID NO:6.

10

15

20

25

30

The present invention especially relates to a HIV-1 Pol mutant such as IA-Pol (SEQ ID NO:4) which comprises a leader peptide, such as the human tPA leader, at the amino terminal portion of the protein, which may effect cellular trafficking and hence, immunogenicity of the expressed protein within the host cell. Any such HIV-1 DNA pol mutant disclosed in the above paragraphs is suitable for fusion downstream of a leader peptide, including but by no means limited to the human tPA leader sequence. Therefore, any such leader peptide-based HIV-1 pol mutant construct may include but is not limited to a mutated DNA molecule which effectively alters the catalytic activity of the RT, RNase and/or IN region of the expressed protein, resulting in at least substantially decreased enzymatic activity one or more of the RT, RNase H and/or IN functions of HIV-1 Pol. In a preferred embodiment of this portion of the invention, a leader peptide/HIV-1 DNA pol construct contains a mutation or mutations within the Pol coding region which effectively abolishes RT, RNase H and IN activity. An especially preferable HIV-1 DNA pol construct is a DNA molecule which contains at least one point mutation which alters the active site and catalytic activity within the RT, RNase H and IN domains of Pol, such that each activity is at least substantially abolished, and preferably totally abolished. Such a HIV-1 Pol mutant will most likely comprise at least one point mutation in or around each catalytic domain responsible for RT, RNase H and IN activity, respectfully. An especially preferred embodiment of this portion of the invention relates to a human tPA leader fused to the IA-Pol protein comprising the nine mutations shown in Table 1. The DNA molecule is disclosed herein as SEQ ID NO:7 and the expressed tPA-IA Pol protein comprises a fusion junction as shown in Figure 3. The complete amino acid sequence of the expressed protein is set forth in SEQ ID NO:8.

The present invention also relates to a substantially purified protein expressed from the DNA polynucleotide vaccines of the present invention, especially the purified

proteins set forth below as SEQ ID NOs: 2, 4, 6, and 8. These purified proteins may be useful as protein-based HIV vaccines.

The present invention also relates to non-codon optimized versions of DNA molecules and associated polynucleotides and associated DNA vaccines which encode the various wild type and modified forms of the HTV Pol protein disclosed herein. Partial or fully codon optimized DNA vaccine expression vector constructs are preferred, but it is within the scope of the present invention to utilize "non-codon optimized" versions of the constructs disclosed herein, especially modified versions of HIV Pol which are shown to promote a substantial cellular immune and humoral immune responses subsequent to host administration.

5

10

15

20

25

30

The DNA backbone of the DNA vaccines of the present invention are preferably DNA plasmid expression vectors. DNA plasmid expression vectors utilized in the present invention include but are not limited to constructs which comprise the cytomegalovirus promoter with the intron A sequence (CMV-intA) and a bovine growth hormone transcription termination sequence. In addition, DNA plasmid vectors of the present invention preferably comprise an antibiotic resistance marker, including but not limited to an ampicillin resistance gene, a neomycin resistance gene or any other pharmaceutically acceptable antibiotic resistance marker. In addition, an appropriate polylinker cloning site and a prokaryotic origin of replication sequence are also preferred. Specific DNA vectors exemplified herein include V1, V1J (SEQ ID NO:13), V1Jneo (SEQ ID NO:14), V1Jns (Figure 1A, SEQ ID NO:15), V1R (SEQ ID NO:26), and any of the aforementioned vectors wherein a nucleotide sequence encoding a leader peptide, preferably the human tPA leader, is fused directly downstream of the CMV-intA promoter, including but not limited to V1Jns-tpa, as shown in Figure 1B and SEQ ID NO:28.

The present invention especially relates to a DNA vaccine and a pharmaceutically active vaccine composition which contains this DNA vaccine, and the use as prophylactic and/or therapeutic vaccine for host immunization, preferably human host immunization, against an HIV infection or to combat an existing HIV condition. These DNA vaccines are represented by codon optimized DNA molecules encoding codon optimized HIV-1 Pol (e.g. SEQ ID NO:2), codon optimized HIV-1 Pol fused to an amino terminal localized leader sequence (e.g. SEQ ID NO:6), and especially preferable, and the essence of the present invention, biologically inactive Pol proteins (IA Pol; e.g., SEQ ID NO:4) devoid of significant PR, RT, RNase or IN

activity associated with wild type Pol and a concomitant construct which contains a leader peptide at the amino terminal region of the IA Pol protein. These constructs are ligated within an appropriate DNA plasmid vector, with or without a nucleotide sequence encoding a functional leader peptide. Preferred DNA vaccines of the present invention comprise codon optimized DNA molecules encoding codon optimized HIV-1 Pol and inactivated version of Pol, ligated in DNA vectors disclosed herein, or any of the aforementioned vectors wherein a nucleotide sequence encoding a leader peptide, preferably the human tPA leader, is fused directly downstream of the CMV-intA promoter, including but not limited to V1Jns-tpa, as shown in Figure 1B and SEQ ID NO:28.

Therefore, the present invention relates to DNA vaccines which include, but are in no way limited to V1Jns-WTPol (comprising the DNA molecule encoding WT Pol, as set forth in SEQ ID NO:2), V1Jns-tPA-WTPol, (comprising the DNA molecule encoding tPA Pol, as set forth in SEQ ID NO:6), V1Jns-IAPol (comprising the DNA molecule encoding IA Pol, as set forth in SEQ ID NO:4), and V1Jns-tPA-IAPol, (comprising the DNA molecule encoding tPA-IA Pol, as set forth in SEQ ID NO:8). Especially preferred are V1Jns-IAPol and V1Jns-tPA-IAPol, as exemplified in Example Section 2.

10

. 20

25

30

The present invention also relates to HIV Pol polynucleotide pharmaceutical products, as well as the production and use thereof, wherein the DNA vaccines are formulated with an adjuvant or adjuvants which may increase immunogenicity of the DNA polynucleotide vaccines of the present invention, namely by promoting an enhanced cellular and/or humoral response subsequent to inoculation. A preferred adjuvant is an aluminum phosphate-based adjuvant or a calcium phosphate based adjuvant, with an aluminum phosphate adjuvant being especially preferred. Another preferred adjuvant is a non-ionic block copolymer, preferably comprising the blocks of polyoxyethylene (POE) and polyoxypropylene (POP) such as a POE-POP-POE block copolymer. These adjuvanted forms comprising the DNA vaccines disclosed herein are useful in increasing cellular responses to DNA vaccination.

As used herein, a DNA vaccine or DNA polynucleotide vaccine is a DNA molecule (i.e., "nucleic acid", "polynucleotide") which contains essential regulatory elements such that upon introduction into a living, vertebrate cell, it is able to direct the cellular machinery to produce translation products encoded by the respective pol

genes of the present invention.

5

10

15

20

25

30

### BRIEF DESCRIPTION OF THE FIGURES

Figure 1A-B shows schematic representation of DNA vaccine expression vectors V1Jns (A) and V1Jns-tPA (B) utilized for HIV-1 pol and HIV-1 modified pol constructs.

Figure 2A-C shows the nucleotide (SEQ ID NO:3) and amino acid sequence (SEQ ID NO:4) of IA-Pol. Underlined codons and amino acids denote mutations, as listed in Table 1.

Figure 3 shows the codon optimized nucleotide and amino acid sequences through the fusion junction of tPA-IA-Pol (contained within SEQ ID NOs: 7 and 8, respectively). The underlined portion represents the NH<sub>2</sub>-terminal region of IA-Pol.

Figure 4 shows generation of a humoral response (measured as the geometric means of anti-RT endpoint titers) from mice immunized with one or two doses of codon optimized V1Jns-IApol and V1Jns-IApol. A portion of mice that received 30 ug of each plasmid was boosted at T=8 wks; sera from all mice were collected at 4 wk post dose 2.

Figure 5 shows the number of IFN-gamma secreting cells per 10e6 cells following stimulation with pools of either CD4<sup>+</sup> (aa641-660, aa731-750) or CD8<sup>+</sup> (aa201-220, aa311-330, aa571-590, aa781-800) specific peptides of splenocytes (pool of 5 spleens/cohort) from control mice and those vaccinated with increasing single dose of codon optimized V1Jns-IApol or 30 ug of codon optimized V1Jns-tpa-IApol (13 wks post dose 1). Mice (n=5) vaccinated with a second dose of 30 ug of either plasmid were analyzed in an Elispot assay at 6 wks post dose 2. Reported are the sums of the number of spots stimulated by each individual CD8<sup>+</sup> peptides because the spots in the wells to which the pool was added are too dense to acquire accurate counts. The CD4<sup>+</sup> cell counts are taken from the responses to the peptide pool. Error bars represent standard deviations for counts from triplicate wells per sample per antigen.

Figure 6A-C shows ELIspot analysis of peripheral blood cells collected from rhesus macaques immunized three times (T=0, 4, 8 wks) with 5 mgs of codon optimized HIV-1 Pol expressing plasmids. Antigen-specific IFN-gamma secretion was stimulated by adding one of two pools consisting of 20-mer peptides derived from vaccine sequence (mpol-1, aa1-420; mpol-2, aa411-850). (A) Frequencies of

spot-forming cells (SFC) as a function of time for 3 monkeys (Tag No. 94R008, 94R013, 94R033) vaccinated with V1Jns-IApol. The reported values are corrected for background responses without peptide restimulation. (B) Frequencies of spot-forming cells (SFC) as a function of time for 3 monkeys (Tag No. 920078, 920073, 94R028) vaccinated with 5mgs of V1Jns-tpa-IApol. (C) ELIspot responses were also measured from a monkey (920072) that did not receive any immunization.

Figure 7A-B show bulk CTL killing from rhesus macaques immunized with codon optimized V1Jns-IApol (A)or codon optimized V1Jns-tpa-IApol (B) at 8 weeks following the third vaccination. Restimulation was performed using recombinant vaccinia virus expressing pol and target cells were prepared by pulsing with the peptide pools, mpol-1 and mpol-2.

Figure 8 shows detection of *in vitro* pol expression from cell lysates of 293 cells transfected with 10 ug of various pol constructs. Bands were detected using antiserum from an HIV-1 seropositive human subject. Equal amounts of total protein were loaded for each lane. The lanes contain the lysates from cells transfected with the following: 1: mock; 2: V1Jns-wt-pol; 3: V1Jns-IApol (codon optimized); 4: V1Jns-tpa-IApol (codon optimized); 5: V1Jns-tpa-pol (codon optimized); 6: V1R-wt-pol (codon optimized); 7: blank; and 8: 80 ng RT.

15

20

25

30

: ;

Figure 9 shows the geometric mean anti-RT titers (GMT) plus the standard errors of the geometric means for cohorts of 5 mice that received one (open circles) or two doses (solid circles) of 1, 10, 100 µg of V1R-wt-pol (codon optimized) or V1Jns-wt-pol. Sera from all animals were collected at 2 weeks post dose 2 (or 7 wks post dose 1) and assayed simultaneously. Statistical analyses were performed to compare cohorts that received the same amount and number of immunization of either plasmids; p values (two-tail) less than 5% are above the bars the connect the correlated cohorts to reflect statistically significant differences.

Figure 10 shows cellular immune responses in BALB/c mice vaccinated i.m. with 1 (pd1) or 2 (pd2) doses of varying amounts of either wt-pol (virus derived) or wt-pol (codon optimized) plasmids. At 3 wks post dose 2, frequencies of IFN-γ-secreting splenocytes are determined from pools of 5 spleens per cohort against mixtures of either CD4<sup>+</sup> peptides (aa21-40, aa411-430, aa531-550, aa641-660, aa731-750, aa771-790) or CD8<sup>+</sup> peptides (aa201-220, aa311-330) at 4 μg/mL final concentration per peptide.

### DETAILED DESCRIPTION OF THE INVENTION

10

15

20

25

30

The present invention relates to synthetic DNA molecules and associated DNA vaccines which elicit CTL and Th cellular immune responses upon administration to the host, including primates and especially humans. An effect of the cellular immune-directed vaccines of the present invention should be a lower transmission rate to previously uninfected individuals and/or reduction in the levels of the viral loads within an infected individual, so as to prolong the asymptomatic phase of HIV-1 infection. In particular, the present invention relates to DNA vaccines which encode various forms of HIV-1 Pol, wherein administration, intracellular delivery and expression of the HIV-1 Pol gene of interest elicits a host CTL and Th response. The preferred synthetic DNA molecules of the present invention encode codon optimized wild type Pol (without Pro activity) and various codon optimized inactivated HIV-1 Pol proteins. The HIV-1 pol constructs disclosed herein are especially preferred for pharmaceutical uses, especially for human administration as a DNA vaccine. The HIV-1 genome employs predominantly uncommon codons compared to highly expressed human genes. Therefore, the pol open reading frame has been synthetically manipulated using optimal codons for human expression. As noted above, a preferred embodiment of the present invention relates to DNA molecules which comprise a HIV-1 pol open reading frame, whether encoding full length pol or a modification or fusion as described herein, wherein the codon usage has been optimized for expression in a mammal, especially a human.

The synthetic *pol* gene disclosed herein comprises the coding sequences for the reverse transcriptase (or RT which consists of a polymerase and RNase H activity) and integrase (IN). The protein sequence is based on that of Hxb2r, a clonal isolate of IIIB; this sequence has been shown to be closest to the consensus clade B sequence with only 16 nonidentical residues out of 848 (Korber, et al., 1998, Human retroviruses and AIDS, Los Alamos National Laboratory, Los Alamos, New Mexico). The skilled artisan will understand after review of this specification that any available HIV-1 or HIV-2 strain provides a potential template for the generation of HIV pol DNA vaccine constructs disclosed herein. It is further noted that the protease gene is excluded from the DNA vaccine constructs of the present invention to insure safety from any residual protease activity in spite of mutational inactivation. The design of the gene sequences for both wild-type (wt-pol) and inactivated pol (IA-pol) incorporates the use of human preferred ("humanized") codons for each amino acid

residue in the sequence in order to maximize *in vivo* mammalian expression (Lathe, 1985, J. Mol. Biol. 183:1-12). As can be discerned by inspecting the codon usage in SEQ ID NOs: 1, 3, 5 and 7, the following codon usage for mammalian optimization is preferred: Met (ATG), Gly (GGC), Lys (AAG), Trp (TGG), Ser (TCC), Arg (AGG), Val (GTG), Pro (CCC), Thr (ACC), Glu (GAG); Leu (CTG), His (CAC), Ile (ATC), Asn (AAC), Cys (TGC), Ala (GCC), Gln (CAG), Phe (TTC) and Tyr (TAC). For an additional discussion relating to mammalian (human) codon optimization, see WO 97/31115 (PCT/US97/02294), which is hereby incorporated by reference. It is intended that the skilled artisan may use alternative versions of codon optimization or may omit this step when generating HIV pol vaccine constructs within the scope of the present invention. Therefore, the present invention also relates to non-codon optimized versions of DNA molecules and associated DNA vaccines which encode the various wild type and modified forms of the HIV Pol protein disclosed herein. However, codon optimization of these constructs is a preferred embodiment of this invention.

10

15

25

30

A particular embodiment of the present invention relates to codon optimized wt-pol DNA constructs (herein, "wt-pol" or "wt-pol (codon optimized))" wherein DNA sequences encoding the protease (PR) activity are deleted, leaving codon optimized "wild type" sequences which encode RT (reverse transcriptase and RNase H activity) and IN integrase activity. A DNA molecule which encodes this protein is disclosed herein as SEQ ID NO:1, the open reading frame being contained from an initiating Met residue at nucleotides 10-12 to a termination codon from nucleotides 2560-2562. SEQ ID NO:1 is as follows:

AGATCTACCA TGGCCCCAT CTCCCCCATT GAGACTGTGC CTGTGAAGCT GAAGCCTGGC
ATGGATGGCC CCAAGGTGAA GCAGTGGCC CTGACTGAGG AGAAGATCAA GGCCCTGGTG
GAAATCTGCA CTGAGATGGA GAAGGAGGGC AAAATCTCCA AGATTGGCCC CGAGAACCCC
TACAACACCC CTGTGTTTGC CATCAAGAAG AAGGACTCCA CCAAGTGGAG GAAGCTGGTG
GACTTCAGGG AGCTGAACAA GAGGACCCAG GACTTCTGGG AGGTGCAGCT GGGCATCCCC
CACCCCGCTG GCCTGAAGAA GAAGAAGTCT GTGACTGTGC TGGATGTGG GGATGCCTAC
TTCTCTGTGC CCCTGGATGA GGACTTCAGG AAGTACACTG CCTTCACCAT CCCCTCCATC
AACAATGAGA CCCCTGGCAT CAGGTACCAG TACAATGTGC TGCCCCAGGG CTGGAAGGGC
TCCCCTGCCA TCTTCCAGTC CTCCATGACC AAGATCCTGG AGCCCTTCAG GAAGCAGAAC
CCTGACATTG TGATCTACCA GTACATGGAT GACCTGTATG TGGGCTCTGA CCTGGAGATT
GGGCAGCACA GGACCAAGAT TGAGGAGCTG AGGCAGCACC TGCTGAGGTG GGGCCTGACC

	ACCCCTGACA	AGAAGCACCA	GAAGGAGCCC	CCCTTCCTGT	GGATGGGCTA	TGAGCTGCAC
	CCCGACAAGT	GGACTGTGCA	GCCCATTGTG	CTGCCTGAGA	AGGACTCCTG	GACTGTGAAT
	GACATCCAGA	AGCTGGTGGG	CAAGCTGAAC	TGGGCCTCCC	AAATCTACCC	TGGCATCAAG
	GTGAGGCAGC	TGTGCAAGCT	GCTGAGGGGC	ACCAAGGCCC	TGACTGAGGT	GATCCCCCTG
5	ACTGAGGAGG	CTGAGCTGGA	GCTGGCTGAG	AACAGGGAGA	TCCTGAAGGA	GCCTGTGCAT
	GGGGTGTACT	ATGACCCCTC	CAAGGACCTG	ATTGCTGAGA	TCCAGAAGCA	GGGCCAGGGC
	CAGTGGACCT	ACCAAATCTA	CCAGGAGCCC	TTCAAGAACC	TGAAGACTGG	CAAGTATGCC
	AGGATGAGGG	GGGCCCACAC	CAATGATGTG	AAGCAGCTGA	CTGAGGCTGT	GCAGAAGATC
	ACCACTGAGT	CCATTGTGAT	CTGGGGCAAG	ACCCCCAAGT	TCAAGCTGCC	CATCCAGAAG
10	GAGACCTGGG	AGACCTGGTG	GACTGAGTAC	TGGCAGGCCA	CCTGGATCCC	TGAGTGGGAG
	TTTGTGAACA	CCCCCCCT	GGTGAAGCTG	TGGTACCAGC	TGGAGAAGGA	GCCCATTGTG
	GGGGCTGAGA	CCTTCTATGT	GGATGGGGCT	GCCAACAGGG	AGACCAAGCT	GGGCAAGGCT .
	GGCTATGTGA	CCAACAGGGG	CAGGCAGAAG	GTGGTGACCC	TGACTGACAC	CACCAACCAG
٠.	AAGACTGAGC	TCCAGGCCAT	CTACCTGGCC	CTCCAGGACT	CTGGCCTGGA	GGTGAACATT
15	GTGACTGACT	CCCAGTATGC	CCTGGGCATC	ATCCAGGCCC	AGCCTGATCA	GTCTGAGTCT
	GAGCTGGTGA	ACCAGATCAT	TGAGCAGCTG	ATCAAGAAGG	AGAAGGTGTA	CCTGGCCTGG
	GTGCCTGCCC	ACAAGGGCAT	TGGGGGCAAT	GAGCAGGTGG	ACAAGCTGGT	GTCTGCTGGC
	ATCAGGAAGG	TGCTGTTCCT	GGATGGCATT	GACAAGGCCC	AGGATGAGCA	TGAGAAGTAC
	CACTCCAACT	GGAGGGCTAT	GGCCTCTGAC	TTCAACCTGC	CCCCTGTGGT	GGCTAAGGAG
20 -	ATTGTGGCCT	CCTGTGACAA	GTGCCAGCTG	AAGGGGGAGG	CCATGCATGG	GCAGGTGGAC
	TGCTCCCCTG	GCATCTGGCA	GCTGGACTGC	ACCCACCTGG	AGGGCAAGGT	GATCCTGGTG
	GCTGTGCATG	TGGCCTCCGG	CTACATTGAG	GCTGAGGTGA	TCCCTGCTGA	GACAGGCCAG
	GAGACTGCCT	ACTTCCTGCT	GAAGCTGGCT	GGCAGGTGGC	CTGTGAAGAC	CATCCACACT
	GACAATGGCT	CCAACTTCAC	TGGGGCCACA	GTGAGGGCTG	CCTGCTGGTG	GGCTGGCATC
25	AAGCAGGAGT	TTGGCATCCC	CTACAACCCC	CAGTCCCAGG	GGGTGGTGGA	GTCCATGAAC
	AAGGAGCTGA	AGAAGATCAT	TGGGCAGGTG	AGGGACCAGG	CTGAGCACCT	GAAGACAGCT
	GTGCAGATGG	CTGTGTTCAT	CCACAACTTC	AAGAGGAAGG	GGGGCATCGG	GGGCTACTCC
	GCTGGGGAGA	GGATTGTGGA	CATCATTGCC	ACAGACATCC	AGACCAAGGA	GCTCCAGAAG
	CAGATCACCA	AGATCCAGAA	CTTCAGGGTG	TACTACAGGG	ACTCCAGGAA	CCCCTGTGG
30	AAGGGCCCTG	CCAAGCTGCT	GTGGAAGGGG	GAGGGGGCTG	TGGTGATCCA	GGACAACTCT
	GACATCAAGG	TGGTGCCCAG	GAGGAAGGCC	AAGATCATCA	GGGACTATGG	CAAGCAGATG
	GCTGGGGATG	ACTGTGTGGC	CTCCAGGCAG	GATGAGGACT	AAAGCCCGGG	CAGATCT (SEQ
	ID NO:1).					

The open reading frame of the wild type pol construct disclosed as SEQ ID NO:1 contains 850 amino acids, disclosed herein as SEQ ID NO:2, as follows: Met Ala Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His Pro Ala Gly Leu Lys Lys Lys Ser Val Thr Val Leu Asp 10 Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val Gly 15 Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu Arg Gln His Leu Leu Arg Trp Gly Leu Thr Thr Pro Asp Lys Lys His Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile 20 Tyr Pro Gly Ile Lys Val Arg Gln Leu Cys Lys Leu Leu Arg Gly Thr Lys Ala Leu Thr Glu Val Ile Pro Leu Thr Glu Glu Ala Glu Leu Glu Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro Val His Gly Val Tyr Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile Gln Lys Gln Gly Gln Gly Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu Lys 25 Thr Gly Lys Tyr Ala Arg Met Arg Gly Ala His Thr Asn Asp Val Lys Gln Leu Thr Glu Ala Val Gln Lys Ile Thr Thr Glu Ser Ile Val Ile Trp Gly Lys Thr Pro Lys Phe Lys Leu Pro Ile Gln Lys Glu Thr Trp Glu Thr Trp Trp Thr Glu Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu Glu 30 Lys Glu Pro Ile Val Gly Ala Glu Thr Phe Tyr Val Asp Gly Ala Ala Asn Arg Glu Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Asn Arg Gly Arg Gln Lys Val Val Thr Leu Thr Asp Thr Thr Asn Gln Lys Thr Glu Leu Gln Ala Ile Tyr Leu Ala Leu Gln Asp Ser Gly Leu Glu Val Asn Ile Val Thr Asp Ser Gln Tyr Ala Leu Gly Ile Ile Gln Ala Gln Pro

Asp Gln Ser Glu Ser Glu Leu Val Asn Gln Ile Ile Glu Gln Leu Ile Lys Lys Glu Lys Val Tyr Leu Ala Trp Val Pro Ala His Lys Gly Ile Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser Ala Gly Ile Arg Lys Val Leu Phe Leu Asp Gly Ile Asp Lys Ala Gln Asp Glu His Glu Lys Tyr His Ser Asn Trp Arg Ala Met Ala Ser Asp Phe Asn Leu Pro Pro Val Val Ala Lys Glu Ile Val Ala Ser Cys Asp Lys Cys Gln Leu Lys Gly Glu Ala Met His Gly Gln Val Asp Cys Ser Pro Gly Ile Trp Gln Leu Asp Cys Thr His Leu Glu Gly Lys Val Ile Leu Val Ala Val His Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile Pro Ala Glu Thr Gly 10 Gln Glu Thr Ala Tyr Phe Leu Leu Lys Leu Ala Gly Arg Trp Pro Val Lys Thr Ile His Thr Asp Asn Gly Ser Asn Phe Thr Gly Ala Thr Val Arg Ala Ala Cys Trp Trp Ala Gly Ile Lys Gln Glu Phe Gly Ile Pro Tyr Asn Pro Gln Ser Gln Gly Val Val Glu Ser Met Asn Lys Glu Leu Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala Glu His Leu Lys Thr Ala Val Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys Gly Gly Ile Gly Gly Tyr Ser Ala Gly Glu Arg Ile Val Asp Ile Ile Ala Thr Asp Ile Gln Thr Lys Glu Leu Gln Lys Gln Ile Thr Lys Ile Gln Asn Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asn Pro Leu Trp Lys Gly Pro Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val Val Ile Gln Asp Asn 20 Ser Asp Ile Lys Val Val Pro Arg Arg Lys Ala Lys Ile Ile Arg Asp Tyr Gly Lys Gln Met Ala Gly Asp Asp Cys Val Ala Ser Arg Gln Asp Glu Asp (SEQ ID NO:2).

The present invention especially relates to a codon optimized HIV-1 DNA pol construct wherein, in addition to deletion of the portion of the wild type sequence encoding the protease activity, a combination of active site residue mutations are introduced which are deleterious to HIV-1 pol (RT-RH-IN) activity of the expressed protein. Therefore, the present invention preferably relates to a HIV-1 DNA pol construct which is devoid of DNA sequences encoding any PR activity, as well as containing a mutation(s) which at least partially, and preferably substantially, abolishes RT, RNase and/or IN activity. One type of HIV-1 pol mutant may include but is not limited to a mutated DNA molecule comprising at least one nucleotide substitution which results in a point mutation which effectively alters an active site within the RT, RNase and/or IN regions of the expressed protein, resulting in at least substantially decreased enzymatic activity for the RT, RNase H and/or IN functions of

30

5

10

15

20

HIV-1 Pol. In a preferred embodiment of this portion of the invention, a HIV-1 DNA pol construct contains a mutation or mutations within the Pol coding region which effectively abolishes RT, RNase H and IN activity. An especially preferable HIV-1 DNA pol construct in a DNA molecule which contains at least one point mutation which alters the active site of the RT, RNase H and IN domains of Pol, such that each activity is at least substantially abolished. Such a HIV-1 Pol mutant will most likely comprise at least one point mutation in or around each catalytic domain responsible for RT, RNase H and IN activity, respectfully. To this end, an especially preferred HIV-1 DNA pol construct is exemplified herein and contains nine codon substitution mutations which results in an inactivated Pol protein (IA Pol: SEQ ID NO:4, Figure 2A-C) which has no PR, RT, RNase or IN activity, wherein three such point mutations reside within each of the RT, RNase and IN catalytic domains. Therefore, an especially preferred exemplification is a DNA molecule which encodes IA-pol, which contains all nine mutations as shown below in Table 1. An additional preferred amino acid residue for substitution is Asp551, localized within the RNase domain of Pol. Any combination of the mutations disclosed herein may suitable and therefore may be utilized as an IA-Pol-based vaccine of the present invention. While addition and deletion mutations are contemplated and within the scope of the invention, the preferred mutation is a point mutation resulting in a substitution of the wild type amino acid with an alternative amino acid residue.

			Table 1	
	wt aa	aa residue	mutant aa	enzyme function
	Asp	112	Ala	RT
25	Asp	187	Ala	RT
	Asp	188	Ala	RT
	Asp	445	Ala	RNase H
30	Glu	480	Ala	RNase H
	Asp	500	Ala	RNase H
	Asp	626	Ala	IN
	Asp	678	Ala	IN
	Glu	714	Ala	IN

It is preferred that point mutations be incorporated into the IApol mutant vaccines of the present invention so as to lessen the possibility of altering epitopes in and around the active site(s) of HIV-1 Pol.

To this end, SEQ ID NO:3 discloses the nucleotide sequence which codes for a codon optimized pol in addition to the nine mutations shown in Table 1, disclosed as follows, and referred to herein as "IApol":

AGATCTACCA TGGCCCCCAT CTCCCCCATT GAGACTGTGC CTGTGAAGCT GAAGCCTGGC ATGGATGGCC CCAAGGTGAA GCAGTGGCCC CTGACTGAGG AGAAGATCAA GGCCCTGGTG GAAATCTGCA CTGAGATGGA GAAGGAGGGC AAAATCTCCA AGATTGGCCC CGAGAACCCC TACAACACCC CTGTGTTTGC CATCAAGAAG AAGGACTCCA CCAAGTGGAG GAAGCTGGTG GACTTCAGGG AGCTGAACAA GAGGACCCAG GACTTCTGGG AGGTGCAGCT GGGCATCCCC CACCCCGCTG GCCTGAAGAA GAAGAAGTCT GTGACTGTGC TGGCTGTGGG GGATGCCTAC TTCTCTGTGC CCCTGGATGA GGACTTCAGG AAGTACACTG CCTTCACCAT CCCCTCCATC AACAATGAGA CCCCTGGCAT CAGGTACCAG TACAATGTGC TGCCCCAGGG CTGGAAGGGC TCCCCTGCCA TCTTCCAGTC CTCCATGACC AAGATCCTGG AGCCCTTCAG GAAGCAGAAC CCTGACATTG TGATCTACCA GTACATGGCT GCCCTGTATG TGGGCTCTGA CCTGGAGATT GGGCAGCACA GGACCAAGAT TGAGGAGCTG AGGCAGCACC TGCTGAGGTG GGGCCTGACC ACCCCTGACA AGAAGCACCA GAAGGAGCCC CCCTTCCTGT GGATGGGCTA TGAGCTGCAC CCCGACAAGT GGACTGTGCA GCCCATTGTG CTGCCTGAGA AGGACTCCTG GACTGTGAAT GACATCCAGA AGCTGGTGGG CAAGCTGAAC TGGGCCTCCC AAATCTACCC TGGCATCAAG GTGAGGCAGC TGTGCAAGCT GCTGAGGGGC ACCAAGGCCC TGACTGAGGT GATCCCCCTG ACTGAGGAGG CTGAGCTGGA GCTGGCTGAG AACAGGGAGA TCCTGAAGGA GCCTGTGCAT GGGGTGTACT ATGACCCCTC CAAGGACCTG ATTGCTGAGA TCCAGAAGCA GGGCCAGGGC CAGTGGACCT ACCAAATCTA CCAGGAGCCC TTCAAGAACC TGAAGACTGG CAAGTATGCC AGGATGAGGG GGGCCCACAC CAATGATGTG AAGCAGCTGA CTGAGGCTGT GCAGAAGATC ACCACTGAGT CCATTGTGAT CTGGGGCAAG ACCCCCAAGT TCAAGCTGCC CATCCAGAAG GAGACCTGGG AGACCTGGTG GACTGAGTAC TGGCAGGCCA CCTGGATCCC TGAGTGGGAG TTTGTGAACA CCCCCCCT GGTGAAGCTG TGGTACCAGC TGGAGAAGGA GCCCATTGTG GGGGCTGAGA CCTTCTATGT GGCTGGGGCT GCCAACAGGG AGACCAAGCT GGGCAAGGCT GGCTATGTGA CCAACAGGGG CAGGCAGAAG GTGGTGACCC TGACTGACAC CACCAACCAG AAGACTGCCC TCCAGGCCAT CTACCTGGCC CTCCAGGACT CTGGCCTGGA GGTGAACATT GTGACTGCCT CCCAGTATGC CCTGGGCATC ATCCAGGCCC AGCCTGATCA GTCTGAGTCT GTGCCTGCCC ACAAGGGCAT TGGGGGCAAT GAGCAGGTGG ACAAGCTGGT GTCTGCTGGC

15

20

25

30

ATCAGGAAGG TGCTGTTCCT GGATGGCATT GACAAGGCCC AGGATGAGCA TGAGAAGTAC CACTCCAACT GGAGGGCTAT GGCCTCTGAC TTCAACCTGC CCCCTGTGGT GGCTAAGGAG ATTGTGGCCT CCTGTGACAA GTGCCAGCTG AAGGGGGGAGG CCATGCATGG GCAGGTGGAC TGCTCCCCTG GCATCTGGCA GCTGGCCTGC ACCCACCTGG AGGGCAAGGT GATCCTGGTG 5 GCTGTGCATG TGGCCTCCGG CTACATTGAG GCTGAGGTGA TCCCTGCTGA GACAGGCCAG GAGACTGCCT ACTTCCTGCT GAAGCTGGCT GGCAGGTGGC CTGTGAAGAC CATCCACACT GCCAATGGCT CCAACTTCAC TGGGGCCACA GTGAGGGCTG CCTGCTGGTG GGCTGGCATC AAGCAGGAGT TTGGCATCCC CTACAACCCC CAGTCCCAGG GGGTGGTGGC CTCCATGAAC AAGGAGCTGA AGAAGATCAT TGGGCAGGTG AGGGACCAGG CTGAGCACCT GAAGACAGCT 10 GTGCAGATGG CTGTGTTCAT CCACAACTTC AAGAGGAAGG GGGCATCGG GGGCTACTCC GCTGGGGAGA GGATTGTGGA CATCATTGCC ACAGACATCC AGACCAAGGA GCTCCAGAAG CAGATCACCA AGATCCAGAA CTTCAGGGTG TACTACAGGG ACTCCAGGAA CCCCCTGTGG AAGGGCCCTG CCAAGCTGCT GTGGAAGGGG GAGGGGGCTG TGGTGATCCA GGACAACTCT GACATCAAGG TGGTGCCCAG GAGGAAGGCC AAGATCATCA GGGACTATGG CAAGCAGATG 15 GCTGGGGATG ACTGTGTGGC CTCCAGGCAG GATGAGGACT AAAGCCCGGG CAGATCT (SEO ID NO:3).

In order to produce the IA-pol DNA vaccine construction, inactivation of the enzymatic functions was achieved by replacing a total of nine active-site residues from the enzyme subunits with alanine side-chains. As shown in Table 1, all residues that comprise the catalytic triad of the polymerase, namely Asp112, Asp187, and Asp188, were substituted with alanine (Ala) residues (Larder, et al., Nature 1987, 327: 716-717; Larder, et al., 1989, Proc. Natl. Acad. Sci. 1989, 86: 4803-4807). Three additional mutations were introduced at Asp445, Glu480 and Asp500 to abolish RNase H activity (Asp551 was left unchanged in this IA Pol construct), with each residue being substituted for an Ala residue, respectively (Davies, et al., 1991, Science 252:, 88-95; Schatz, et al., 1989, FEBS Lett. 257: 311-314; Mizrahi, et al., 1990, Nucl. Acids. Res. 18: pp. 5359-5353). HIV pol integrase function was abolished through three mutations at Asp626, Asp678 and Glu714. Again, each of these residues has been substituted with an Ala residue (Wiskerchen, et al., 1995, J. Virol. 69: 376-386; Leavitt, et al., 1993, J. Biol. Chem. 268: 2113-2119). Amino acid residue Pro3 of SEQ ID NO:4 marks the start of the RT gene. The complete amino acid sequence of IA-Pol is disclosed herein as SEQ ID NO:4, as follows: Met Ala Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu Lys

20

25

30

Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His Pro Ala Gly Leu Lys Lys Lys Ser Val Thr Val Leu Ala Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Ala Ala Leu Tyr Val Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu Arg Gln His Leu Leu Arg Trp Gly Leu Thr Thr Pro Asp Lys Lys His Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Pro Gly Ile Lys Val Arg Gln Leu Cys Lys Leu Leu Arg Gly Thr Lys Ala Leu Thr Glu Val Ile Pro Leu Thr Glu Glu Ala Glu Leu Glu Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro Val His Gly Val Tyr Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile Gln Lys Gln Gly Gln Gly Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu Lys Thr Gly Lys Tyr Ala Arg Met Arg Gly Ala His Thr Asn Asp Val Lys Gln Leu Thr Glu Ala Val Gln Lys Ile Thr Thr Glu Ser Ile Val Ile Trp Gly Lys Thr Pro Lys Phe Lys Leu Pro Ile Gln Lys Glu Thr Trp Glu Thr Trp Trp Thr Glu Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp 25 Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu Glu Lys Glu Pro Ile Val Gly Ala Glu Thr Phe Tyr Val Ala Gly Ala Ala Asn Arg Glu Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Asn Arg Gly Arg Gln Lys Val Val Thr Leu Thr Asp Thr Thr Asn Gln Lys Thr Ala Leu Gln Ala Ile Tyr Leu Ala Leu Gln Asp Ser Gly Leu Glu Val Asn Ile Val Thr Ala Ser Gln Tyr Ala Leu Gly Ile Ile Gln Ala Gln Pro Asp Gln Ser Glu Ser Glu Leu Val Asn Gln Ile Ile Glu Gln Leu Ile Lys Lys Glu Lys Val Tyr Leu Ala Trp Val Pro Ala His Lys Gly Ile Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser Ala Gly Ile Arg Lys Val Leu Phe Leu Asp Gly Ile Asp Lys Ala Gln Asp Glu His Glu Lys

Tyr His Ser Asn Trp Arg Ala Met Ala Ser Asp Phe Asn Leu Pro Pro Val Val Ala Lys Glu Ile Val Ala Ser Cys Asp Lys Cys Gln Leu Lys Gly Glu Ala Met His Gly Gln Val Asp Cys Ser Pro Gly Ile Trp Gln Leu Ala Cys Thr His Leu Glu Gly Lys Val Ile Leu Val Ala Val His 5 Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile Pro Ala Glu Thr Gly Gln Glu Thr Ala Tyr Phe Leu Leu Lys Leu Ala Gly Arg Trp Pro Val Lys Thr Ile His Thr Ala Asn Gly Ser Asn Phe Thr Gly Ala Thr Val Arg Ala Ala Cys Trp Trp Ala Gly Ile Lys Gln Glu Phe Gly Ile Pro Tyr Asn Pro Gln Ser Gln Gly Val Val Ala Ser Met Asn Lys Glu Leu 10 Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala Glu His Leu Lys Thr Ala Val Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys Gly Gly Ile Gly Gly Tyr Ser Ala Gly Glu Arg Ile Val Asp Ile Ile Ala Thr Asp Ile Gln Thr Lys Glu Leu Gln Lys Gln Ile Thr Lys Ile Gln Asn Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asn Pro Leu Trp Lys Gly Pro 15 Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val Val Ile Gln Asp Asn Ser Asp Ile Lys Val Val Pro Arg Arg Lys Ala Lys Ile Ile Arg Asp Tyr Gly Lys Gln Met Ala Gly Asp Asp Cys Val Ala Ser Arg Gln Asp Glu Asp (SEQ ID NO:4).

As noted above, it will be understood that any combination of the mutations disclosed above may be suitable and therefore be utilized as an IA-pol-based vaccine of the present invention. For example, it may be possible to mutate only 2 of the 3 residues within the respective reverse transcriptase, RNase H, and integrase coding regions while still abolishing these enzymatic activities. However, the IA-pol construct described above and disclosed as SEQ ID NO:3, as well as the expressed protein (SEQ ID NO:4) is preferred. It is also preferred that at least one mutation be present in each of the three catalytic domains.

20

25

30

Another aspect of the present invention is to generate codon optimized HIV-1 Pol-based vaccine constructions which comprise a eukaryotic trafficking signal peptide such as from tPA (tissue-type plasminogen activator) or by a leader peptide such as is found in highly expressed mammalian proteins such as immunoglobulin leader peptides. Any functional leader peptide may be tested for efficacy. However, a preferred embodiment of the present invention is to provide for HIV-1 Pol mutant vaccine constructions as disclosed herein which also comprise a leader peptide, preferably a leader peptide from human tPA. In other words, a codon optimized

HIV-1 Pol mutant such as IA-Pol (SEQ ID NO:4) may also comprise a leader peptide at the amino terminal portion of the protein, which may effect cellular trafficking and hence, immunogenicity of the expressed protein within the host cell. As shown in Figure 1A-B for the DNA vector V1Jns, a DNA vector which may be utilized to practice the present invention may be modified by known recombinant DNA methodology to contain a leader signal peptide of interest, such that downstream cloning of the modified HIV-1 protein of interest results in a nucleotide sequence which encodes a modified HIV-1 tPA/Pol protein. In the alternative, as noted above, insertion of a nucleotide sequence which encodes a leader peptide may be inserted into a DNA vector housing the open reading frame for the Pol protein of interest. Regardless of the cloning strategy, the end result is a polynucleotide vaccine which comprises vector components for effective gene expression in conjunction with nucleotide sequences which encode a modified HIV-1 Pol protein of interest, including but not limited to a HIV-1 Pol protein which contains a leader peptide. The amino acid sequence of the human tPA leader utilized herein is as follows: MDAMKRGLCCVLLLCGAVFVSPSEISS (SEQ ID NO:28). Therefore, another aspect of the present invention is to generate HIV-1 Pol-based vaccine constructions which comprise a eukaryotic trafficking signal peptide such as from tPA. To this end, the present invention relates to a DNA molecule which encodes a codon optimized wt-pol DNA construct wherein the protease (PR) activity is deleted and a human tPA leader sequence is fused to the 5' end of the coding region. A DNA molecule which encodes this protein is disclosed herein as SEQ ID NO:5, the open reading frame disclosed herein as SEQ ID NO:6.

10

15

20

25

30

To this end, the present invention relates to a DNA molecule which encodes a codon optimized wt-pol DNA construct wherein the protease (PR) activity is deleted and a human tPA leader sequence is fused to the 5' end of the coding region (herein, "tPA-wt-pol"). A DNA molecule which encodes this protein is disclosed herein as SEQ ID NO:5, the open reading frame being contained from an initiating Met residue at nucleotides 8-10 to a termination codon from nucleotides 2633-2635. SEQ ID NO:5 is as follows:

GATCACCATG GATGCAATGA AGAGAGGGCT CTGCTGTGTG CTGCTGCTGT GTGGAGCAGT
CTTCGTTTCG CCCAGCGAGA TCTCCGCCCC CATCTCCCCC ATTGAGACTG TGCCTGTGAA
GCTGAAGCCT GGCATGGATG GCCCCAAGGT GAAGCAGTGG CCCCTGACTG AGGAGAAGAT
CAAGGCCCTG GTGGAAATCT GCACTGAGAT GGAGAAGGAG GGCAAAATCT CCAAGATTGG

	CCCCGAGAAC	CCCTACAACA	CCCCTGTGTT	TGCCATCAAG	AAGAAGGACT	CCACCAAGTG
	GAGGAAGCTG	GTGGACTTCA	GGGAGCTGAA	CAAGAGGACC	CAGGACTTCT	GGGAGGTGCA
	GCTGGGCATC	CCCCACCCCG	CTGGCCTGAA	GAAGAAGAAG	TCTGTGACTG	TGCTGGATGT
	GGGGGATGCC	TACTTCTCTG	TGCCCCTGGA	TGAGGACTTC	AGGAAGTACA	CTGCCTTCAC
5	CATCCCCTCC	ATCAACAATG	AGACCCCTGG	CATCAGGTAC	CAGTACAATG	TGCTGCCCCA
	GGGCTGGAAG	GGCTCCCCTG	CCATCTTCCA	GTCCTCCATG	ACCAAGATCC	TGGAGCCCTT
	CAGGAAGCAG	AACCCTGACA	TTGTGATCTA	CCAGTACATG	GATGACCTGT	ATGTGGGCTC
	TGACCTGGAG	ATTGGGCAGC	ACAGGACCAA	GATTGAGGAG	CTGAGGCAGC	ACCTGCTGAG
	GTGGGGCCTG	ACCACCCCTG	ACAAGAAGCA	CCAGAAGGAG	CCCCCCTTCC	TGTGGATGGG
10	CTATGAGCTG	CACCCCGACA	AGTGGACTGT	GCAGCCCATT	GTGCTGCCTG	AGAAGGACTC
	CTGGACTGTG	AATGACATCC	AGAAGCTGGT	GGGCAAGCTG	AACTGGGCCT	CCCAAATCTA
	CCCTGGCATC	AAGGTGAGGC	AGCTGTGCAA	GCTGCTGAGG	GGCACCAAGG	CCCTGACTGA
	GGTGATCCCC	CTGACTGAGG	AGGCTGAGCT	GGAGCTGGCT	GAGAACAGGG	AGATCCTGAA
	GGAGCCTGTG	CATGGGGTGT	ACTATGACCC	CTCCAAGGAC	CTGATTGCTG	AGATCCAGAA
15	GCAGGGCCAG	GGCCAGTGGA	CCTACCAAAT	CTACCAGGAG	CCCTTCAAGA	ACCTGAAGAC
	TGGCAAGTAT	GCCAGGATGA	GGGGGCCCA	CACCAATGAT	GTGAAGCAGC	TGACTGAGGC
	TGTGCAGAAG	ATCACCACTG	AGTCCATTGT	GATCTGGGGC	AAGACCCCCA	AGTTCAAGCT
	GCCCATCCAG	AAGGAGACCT	GGGAGACCTG	GTGGACTGAG	TACTGGCAGG	CCACCTGGAT
	CCCTGAGTGG	GAGTTTGTGA	ACACCCCCC	CCTGGTGAAG	CTGTGGTACC	AGCTGGAGAA
20	GGAGCCCATT	GTGGGGGCTG	AGACCTTCTA	TGTGGATGGG	GCTGCCAACA	GGGAGACCAA
	GCTGGGCAAG	GCTGGCTATG	TGACCAACAG	GGGCAGGCAG	AAGGTGGTGA	CCCTGACTGA
	CACCACCAAC	CAGAAGACTG	AGCTCCAGGC	CATCTACCTG	GCCCTCCAGG	ACTCTGGCCT
	GGAGGTGAAC	ATTGTGACTG	ACTCCCAGTA	TGCCCTGGGC	ATCATCCAGG	CCCAGCCTGA
	TCAGTCTGAG	TCTGAGCTGG	TGAACCAGAT	CATTGAGCAG	CTGATCAAGA	AGGAGAAGGT
25	GTACCTGGCC	TGGGTGCCTG	CCCACAAGGG	CATTGGGGGC	AATGAGCAGG	TGGACAAGCT
	GGTGTCTGCT	GGCATCAGGA	AGGTGCTGTT	CCTGGATGGC	ATTGACAAGG	CCCAGGATGA
	GCATGAGAAG	TACCACTCCA	ACTGGAGGGC	TATGGCCTCT	GACTTCAACC	TGCCCCCTGT
	GGTGGCTAAG	GAGATTGTGG	CCTCCTGTGA	CAAGTGCCAG	CTGAAGGGGG	AGGCCATGCA
	TGGGCAGGTG	GACTGCTCCC	CTGGCATCTG	GCAGCTGGAC	TGCACCCACC	TGGAGGGCAA
30	GGTGATCCTG	GTGGCTGTGC	ATGTGGCCTC	CGGCTACATT	GAGGCTGAGG	TGATCCCTGC
	TGAGACAGGC	CAGGAGACTG	CCTACTTCCT	GCTGAAGCTG	GCTGGCAGGT	GGCCTGTGAA
	GACCATCCAC	ACTGACAATG	GCTCCAACTT	CACTGGGGCC	ACAGTGAGGG	CTGCCTGCTG
	GTGGGCTGGC	ATCAAGCAGG	AGTTTGGCAT	CCCCTACAAC	CCCCAGTCCC	AGGGGGTGGT
	GGAGTCCATG	AACAAGGAGC	TGAAGAAGAT	CATTGGGCAG	GTGAGGGACC	AGGCTGAGCA

CCTGAAGACA GCTGTGCAGA TGGCTGTGTT CATCCACAAC TTCAAGAGGA AGGGGGGCAT
CGGGGGCTAC TCCGCTGGGG AGAGGATTGT GGACATCATT GCCACAGACA TCCAGACCAA
GGAGCTCCAG AAGCAGATCA CCAAGATCCA GAACTTCAGG GTGTACTACA GGGACTCCAG
GAACCCCCTG TGGAAGGGCC CTGCCAAGCT GCTGTGGAAG GGGGAGGGGG CTGTGGTGAT
CCAGGACAAC TCTGACATCA AGGTGGTGCC CAGGAGGAAG GCCAAGATCA TCAGGGACTA
TGGCAAGCAG ATGGCTGGGG ATGACTGTGT GGCCTCCAGG CAGGATGAG ACTAAAGCCC
GGGCAGATCT (SEQ ID NO:5).

The open reading frame of the wild type tPA-pol construct disclosed as SEQ ID NO:5 contains 875 amino acids, disclosed herein as SEQ ID NO:6, as follows:

10 Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Cys Gly Ala Val Phe Val Ser Pro Ser Glu Ile Ser Ala Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser Lys Ile Gly Pro Glu 15 Asn Pro Tyr Asn Thr Pro Val Phe Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg Lys Tyr Thr Ala Phe Thr Ile Pro 20 Ser Ile Asn Asn Glu Thr Pro Gly Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu Arg Gln His Leu Leu Arg Trp Gly 25 Leu Thr Thr Pro Asp Lys Lys His Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Pro Gly Ile Lys Val Arg Gln Leu Cys Lys Leu Leu Arg Gly Thr Lys Ala Leu Thr Glu Val Ile 30 Pro Leu Thr Glu Glu Ala Glu Leu Glu Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro Val His Gly Val Tyr Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile Gln Lys Gln Gly Gln Gly Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu Lys Thr Gly Lys Tyr Ala Arg Met Arg Gly Ala His Thr Asn Asp Val Lys Gln Leu Thr Glu Ala Val Gln

Lys Ile Thr Thr Glu Ser Ile Val Ile Trp Gly Lys Thr Pro Lys Phe Lys Leu Pro Ile Gln Lys Glu Thr Trp Glu Thr Trp Trp Thr Glu Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu Glu Lys Glu Pro Ile Val Gly Ala Glu Thr Phe Tyr Val Asp Gly Ala Ala Asn Arg Glu Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Asn Arg Gly Arg Gln Lys Val Val Thr Leu Thr Asp Thr Thr Asn Gln Lys Thr Glu Leu Gln Ala Ile Tyr Leu Ala Leu Gln Asp Ser Gly Leu Glu Val Asn Ile Val Thr Asp Ser Gln Tyr Ala Leu Gly Ile Ile Gln Ala Gln Pro Asp Gln Ser Glu Ser Glu Leu 10 Val Asn Gln Ile Ile Glu Gln Leu Ile Lys Lys Glu Lys Val Tyr Leu Ala Trp Val Pro Ala His Lys Gly Ile Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser Ala Gly Ile Arg Lys Val Leu Phe Leu Asp Gly Ile Asp Lys Ala Gln Asp Glu His Glu Lys Tyr His Ser Asn Trp Arg Ala Met Ala Ser Asp Phe Asn Leu Pro Pro Val Val Ala Lys Glu Ile Val Ala Ser Cys Asp Lys Cys Gln Leu Lys Gly Glu Ala Met His Gly Gln Val Asp Cys Ser Pro Gly Ile Trp Gln Leu Asp Cys Thr His Leu Glu Gly Lys Val Ile Leu Val Ala Val His Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile Pro Ala Glu Thr Gly Gln Glu Thr Ala Tyr Phe Leu Leu Lys Leu Ala Gly Arg Trp Pro Val Lys Thr Ile His Thr Asp Asn Gly Ser Asn Phe Thr Gly Ala Thr Val Arg Ala Ala Cys Trp Trp Ala Gly Ile Lys Gln Glu Phe Gly Ile Pro Tyr Asn Pro Gln Ser Gln Gly Val Val Glu Ser Met Asn Lys Glu Leu Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala Glu His Leu Lys Thr Ala Val Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys Gly Gly Ile Gly Gly Tyr Ser Ala Gly 25 Glu Arg Ile Val Asp Ile Ile Ala Thr Asp Ile Gln Thr Lys Glu Leu Gln Lys Gln Ile Thr Lys Ile Gln Asn Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asn Pro Leu Trp Lys Gly Pro Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val Val Ile Gln Asp Asn Ser Asp Ile Lys Val Val Pro Arg Arg Lys Ala Lys Ile Ile Arg Asp Tyr Gly Lys Gln Met Ala Gly 30 Asp Asp Cys Val Ala Ser Arg Gln Asp Glu Asp (SEQ ID NO:6).

The present invention also relates to a codon optimized HIV-1 Pol mutant such as IA-Pol (SEQ ID NO:4) which comprises a leader peptide at the amino terminal portion of the protein, which may effect cellular trafficking and hence, immunogenicity of the expressed protein within the host cell. Any such HIV-1 DNA

pol mutant disclosed in the above paragraphs is suitable for fusion downstream of a leader peptide, such as a leader peptide including but not limited to the human tPA leader sequence. Therefore, any such leader peptide-based HIV-1 pol mutant construct may include but is not limited to a mutated DNA molecule which effectively alters the catalytic activity of the RT, RNase and/or IN region of the expressed protein, 5 resulting in at least substantially decreased enzymatic activity one or more of the RT, RNase H and/or IN functions of HIV-1 Pol. In a preferred embodiment of this portion of the invention, a leader peptide/HIV-1 DNA pol construct contains a mutation or mutations within the Pol coding region which effectively abolishes RT, RNase H and 10 IN activity. An especially preferable HIV-1 DNA pol construct is a DNA molecule which contains at least one point mutation which alters the active site and catalytic activity within the RT, RNase H and IN domains of Pol, such that each activity is at least substantially abolished, and preferably totally abolished. Such a HIV-1 Pol mutant will most likely comprise at least one point mutation in or around each-15 catalytic domain responsible for RT, RNase H and IN activity, respectfully. An especially preferred embodiment of this portion of the invention relates to a human tPA leader fused to the IA-Pol protein comprising the nine mutations shown in Table 1. The DNA molecule is disclosed herein as SEQ ID NO:7 and the expressed tPA-IA Pol protein comprises a fusion junction as shown in Figure 3. The complete amino 20 acid sequence of the expressed protein is set forth in SEQ ID NO:8. To this end, SEQ ID NO:7 discloses the nucleotide sequence which codes for a human tPA leader fused to the IA Pol protein comprising the nine mutations shown in Table 1 (herein, "tPAopt-IApol"). The open reading frame begins with the initiating Met (nucleotides 8-10) and terminates with a "TAA" codon at nucleotides 2633-2635. The nucleotide 25 sequence encoding tPA-IAPol is also disclosed as follows: GATCACCATG GATGCAATGA AGAGAGGGCT CTGCTGTGTG CTGCTGCTGT GTGGAGCAGT CTTCGTTTCG CCCAGCGAGA TCTCCGCCCC CATCTCCCCC ATTGAGACTG TGCCTGTGAA GCTGAAGCCT GGCATGGATG GCCCCAAGGT GAAGCAGTGG CCCCTGACTG AGGAGAAGAT CAAGGCCCTG GTGGAAATCT GCACTGAGAT GGAGAAGGAG GGCAAAATCT CCAAGATTGG 30 CCCCGAGAAC CCCTACAACA CCCCTGTGTT TGCCATCAAG AAGAAGGACT CCACCAAGTG GAGGAAGCTG GTGGACTTCA GGGAGCTGAA CAAGAGGACC CAGGACTTCT GGGAGGTGCA GCTGGGCATC CCCCACCCG CTGGCCTGAA GAAGAAGAG TCTGTGACTG TGCTGGCTGT GGGGGATGCC TACTTCTCTG TGCCCCTGGA TGAGGACTTC AGGAAGTACA CTGCCTTCAC CATCCCCTCC ATCAACAATG AGACCCCTGG CATCAGGTAC CAGTACAATG TGCTGCCCCA

	GGGCTGGAAG	GGCTCCCCTG	CCATCTTCCA	GTCCTCCATG	ACCAAGATCC	TGGAGCCCTT
	CAGGAAGCAG	AACCCTGACA	TTGTGATCTA	CCAGTACATG	GCTGCCCTGT	ATGTGGGCTC
	TGACCTGGAG	ATTGGGCAGC	ACAGGACCAA	GATTGAGGAG	CTGAGGCAGC	ACCTGCTGAG
	GTGGGGCCTG	ACCACCCCTG	ACAAGAAGCA	CCAGAAGGAG	CCCCCTTCC	TGTGGATGGG
5	CTATGAGCTG	CACCCGACA	AGTGGACTGT	GCAGCCCATT	GTGCTGCCTG	AGAAGGACTC
	CTGGACTGTG	AATGACATCC	AGAAGCTGGT	GGGCAAGCTG	AACTGGGCCT	CCCAAATCTA
	CCCTGGCATC	AAGGTGAGGC	AGCTGTGCAA	GCTGCTGAGG	GGCACCAAGG	CCCTGACTGA
	GGTGATCCCC	CTGACTGAGG	AGGCTGAGCT	GGAGCTGGCT	GAGAACAGGG	AGATCCTGAA
	GGAGCCTGTG	CATGGGGTGT	ACTATGACCC	CTCCAAGGAC	CTGATTGCTG	AGATCCAGAA
10	GCAGGGCCAG	GGCCAGTGGA	CCTACCAAAT	CTACCAGGAG	CCCTTCAAGA	ACCTGAAGAC
	TGGCAAGTAT	GCCAGGATGA	GGGGGCCCA	CACCAATGAT	GTGAAGCAGC	TGACTGAGGC
	TGTGCAGAAG	ATCACCACTG	AGTCCATTGT	GATCTGGGGC	AAGACCCCCA	AGTTCAAGCT
	GCCCATCCAG	AAGGAGACCT	GGGAGACCTG	GTGGACTGAG	TACTGGCAGG	CCACCTGGAT
	CCCTGAGTGG	GAGTTTGTGA	ACACCCCCC	CCTGGTGAAG	CTGTGGTACC	AGCTGGAGAA
15	GGAGCCCATT	GTGGGGGCTG	AGACCTTCTA	TGTGGCTGGG	GCTGCCAACA	GGGAGACCAA
	GCTGGGCAAG	GCTGGCTATG	TGACCAACAG	GGGCAGGCAG	AAGGTGGTGA	CCCTGACTGA
	CACCACCAAC	CAGAAGACTG	CCCTCCAGGC	CATCTACCTG	GCCCTCCAGG	ACTCTGGCCT
	GGAGGTGAAC	ATTGTGACTG	CCTCCCAGTA	TGCCCTGGGC	ATCATCCAGG	CCCAGCCTGA
	TCAGTCTGAG	TCTGAGCTGG	TGAACCAGAT	CATTGAGCAG	CTGATCAAGA	AGGAGAAGGT
20	GTACCTGGCC	TGGGTGCCTG	CCCACAAGGG	CATTGGGGGC	AATGAGCAGG	TGGACAAGCT
	GGTGTCTGCT	GGCATCAGGA	AGGTGCTGTT	CCTGGATGGC	ATTGACAAGG	CCCAGGATGA
	GCATGAGAAG	TACCACTCCA	ACTGGAGGGC	TATGGCCTCT	GACTTCAACC	TGCCCCCTGT
•	GGTGGCTAAG	GAGATTGTGG	CCTCCTGTGA	CAAGTGCCAG	CTGAAGGGGG	AGGCCATGCA
	TGGGCAGGTG	GACTGCTCCC	CTGGCATCTG	GCAGCTGGCC	TGCACCCACC	TGGAGGGCAA
25	GGTGATCCTG	GTGGCTGTGC	ATGTGGCCTC	CGGCTACATT	GAGGCTGAGG	TGATCCCTGC
		CAGGAGACTG				
	GACCATCCAC	ACTGCCAATG	GCTCCAACTT	CACTGGGGCC	ACAGTGAGGG	CTGCCTGCTG
	GTGGGCTGGC	ATCAAGCAGG	AGTTTGGCAT	CCCCTACAAC	CCCCAGTCCC	AGGGGGTGGT
	GGCCTCCATG	AACAAGGAGC	TGAAGAAGAT	CATTGGGCAG	GTGAGGGACC	AGGCTGAGCA
30	CCTGAAGACA	GCTGTGCAGA	TGGCTGTGTT	CATCCACAAC	TTCAAGAGGA	AGGGGGCAT
	CGGGGGCTAC	TCCGCTGGGG	AGAGGATTGT	GGACATCATT	GCCACAGACA	TCCAGACCAA
	GGAGCTCCAG	AAGCAGATCA	CCAAGATCCA	GAACTTCAGG	GTGTACTACA	GGGACTCCAG
	GAACCCCCTG	TGGAAGGGCC	CTGCCAAGCT	GCTGTGGAAG	GGGGAGGGG	CTGTGGTGAT
	CCAGGACAAC	TCTGACATCA	AGGTGGTGCC	CAGGAGGAAG	GCCAAGATCA	TCAGGGACTA

TGGCAAGCAG ATGGCTGGGG ATGACTGTGT GGCCTCCAGG CAGGATGAGG ACTAAAGCCC GGGCAGATCT (SEQ ID NO:7).

The open reading frame of the tPA-IA-pol construct disclosed as SEQ ID NO:7 contains 875 amino acids, disclosed herein as tPA-IA-Pol and SEQ ID NO:8, as

follows:

Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Cys Gly Ala Val Phe Val Ser Pro Ser Glu Ile Ser Ala Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu Lys Ile Lys Ala Leu Val Glu Ile 10 Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His Pro Ala Gly Leu Lys Lys Lys Ser Val Thr Val Leu Ala Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Ala Ala Leu Tyr Val Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu Arg Gln His Leu Leu Arg Trp Gly 20 Leu Thr Thr Pro Asp Lys Lys His Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Pro Gly Ile Lys Val Arg 25 Gln Leu Cys Lys Leu Leu Arg Gly Thr Lys Ala Leu Thr Glu Val Ile Pro Leu Thr Glu Glu Ala Glu Leu Glu Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro Val His Gly Val Tyr Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile Gln Lys Gln Gly Gln Gly Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu Lys Thr Gly Lys Tyr Ala Arg Met Arg Gly Ala His Thr Asn Asp Val Lys Gln Leu Thr Glu Ala Val Gln Lys Ile Thr Thr Glu Ser Ile Val Ile Trp Gly Lys Thr Pro Lys Phe Lys Leu Pro Ile Gln Lys Glu Thr Trp Glu Thr Trp Trp Thr Glu Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu Glu Lys Glu Pro Ile Val Gly Ala

Glu Thr Phe Tyr Val Ala Gly Ala Ala Asn Arg Glu Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Asn Arg Gly Arg Gln Lys Val Val Thr Leu Thr Asp Thr Thr Asn Gln Lys Thr Ala Leu Gln Ala Ile Tyr Leu Ala Leu Gln Asp Ser Gly Leu Glu Val Asn Ile Val Thr Ala Ser Gln Tyr Ala Leu Gly Ile Ile Gln Ala Gln Pro Asp Gln Ser Glu Ser Glu Leu Val Asn Gln Ile Ile Glu Gln Leu Ile Lys Lys Glu Lys Val Tyr Leu Ala Trp Val Pro Ala His Lys Gly Ile Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser Ala Gly Ile Arg Lys Val Leu Phe Leu Asp Gly Ile Asp Lys Ala Gln Asp Glu His Glu Lys Tyr His Ser Asn Trp Arg Ala Met Ala Ser Asp Phe Asn Leu Pro Pro Val Val Ala Lys Glu Ile Val Ala Ser Cys Asp Lys Cys Gln Leu Lys Gly Glu Ala Met His Gly Gln Val Asp Cys Ser Pro Gly Ile Trp Gln Leu Ala Cys Thr His Leu Glu Gly Lys Val Ile Leu Val Ala Val His Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile Pro Ala Glu Thr Gly Gln Glu Thr Ala Tyr Phe Leu Leu Lys Leu Ala Gly Arg Trp Pro Val Lys Thr Ile His Thr Ala Asn Gly Ser Asn Phe Thr Gly Ala Thr Val Arg Ala Ala Cys Trp Trp Ala Gly Ile Lys Gln Glu Phe Gly Ile Pro Tyr Asn Pro Gln Ser Gln Gly Val Val Ala Ser Met Asn Lys Glu Leu Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala Glu His Leu Lys Thr Ala Val Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys Gly Gly Ile Gly Gly Tyr Ser Ala Gly Glu Arg Ile Val Asp Ile Ile Ala Thr Asp Ile Gln Thr Lys Glu Leu Gln Lys Gln Ile Thr Lys Ile Gln Asn Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asn Pro Leu Trp Lys Gly Pro Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val Val Ile Gln Asp Asn Ser Asp Ile Lys Val Val Pro Arg Arg Lys Ala Lys Ile Ile Arg Asp Tyr Gly Lys Gln Met Ala Gly Asp Asp Cys Val Ala Ser Arg Gln Asp Glu Asp (SEQ ID NO:8).

The present invention also relates to a substantially purified protein expressed from the DNA polynucleotide vaccines of the present invention, especially the purified proteins set forth below as SEQ ID NOs: 2, 4, 6, and 8. These purified proteins may be useful as protein-based HIV vaccines.

30

The DNA backbone of the DNA vaccines of the present invention are preferably DNA plasmid expression vectors. DNA plasmid expression vectors are well known in the art and the present DNA vector vaccines may be comprised of any such expression backbone which contains at least a promoter for RNA polymerase

transcription, and a transcriptional terminator 3' to the HIV pol coding sequence. In one preferred embodiment, the promoter is the Rous sarcoma virus (RSV) long terminal repeat (LTR) which is a strong transcriptional promoter. A more preferred promoter is the cytomegalovirus promoter with the intron A sequence (CMV-intA). A preferred transcriptional terminator is the bovine growth hormone terminator. In addition, to assist in large scale preparation of an HIV pol DNA vector vaccine, an antibiotic resistance marker is also preferably included in the expression vector. Ampicillin resistance genes, neomycin resistance genes or any other pharmaceutically acceptable antibiotic resistance marker may be used. In a preferred embodiment of this invention, the antibiotic resistance gene encodes a gene product for neomycin resistance. Further, to aid in the high level production of the pharmaceutical by fermentation in prokaryotic organisms, it is advantageous for the vector to contain an origin of replication and be of high copy number. Any of a number of commercially available prokaryotic cloning vectors provide these benefits. In a preferred embodiment of this invention, these functionalities are provided by the commercially available vectors known as pUC. It is desirable to remove non-essential DNA sequences. Thus, the lacZ and lacI coding sequences of pUC are removed in one embodiment of the invention.

5

10

15

20

25

30

DNA expression vectors which exemplify but in no way limit the present invention are disclosed in PCT International Application No. PCT/US94/02751, International Publication No. WO 94/21797, hereby incorporated by reference. A first DNA expression vector is the expression vector pnRSV, wherein the rous sarcoma virus (RSV) long terminal repeat (LTR) is used as the promoter. A second embodiment relates to plasmid V1, a mutated pBR322 vector into which the CMV promoter and the BGH transcriptional terminator is cloned. Another embodiment regarding DNA vector backbones relates to plasmid V1J. Plasmid V1J is derived from plasmid V1 and removes promoter and transcription termination elements in order to place them within a more defined context, create a more compact vector, and to improve plasmid purification yields. Therefore, V1J also contains the CMVintA promoter and (BGH) transcription termination elements which control the expression of the HIV pol-based genes disclosed herein. The backbone of V1J is provided by pUC18. It is known to produce high yields of plasmid, is well-characterized by sequence and function, and is of minimum size. The entire lac operon was removed and the remaining plasmid was purified from an agarose electrophoresis gel.

blunt-ended with the T4 DNA polymerase, treated with calf intestinal alkaline phosphatase, and ligated to the CMVintA/BGH element. In a preferred DNA expression vector, the ampicillin resistance gene is removed from VIJ and replaced with a neomycin resistance gene, to generate VIJneo. An especially preferred DNA expression vector is V1Jns, which is the same as V1J except that a unique Sfi1 restriction site has been engineered into the single Kpn1 site at position 2114 of V1Jneo. The incidence of Sfi1 sites in human genomic DNA is very low (approximately 1 site per 100,000 bases). Thus, this vector allows careful monitoring for expression vector integration into host DNA, simply by Sfi1 digestion of extracted genomic DNA. Yet another preferred DNA expression vector used as the backbone to the HIV-1 pol-based DNA vaccines of the present invention is V1R. In this vector, as much non-essential DNA as possible is "trimmed" from the vector to produce a highly compact vector. This vector is a derivative of V1Jns. This vector allows larger inserts to be used, with less concern that undesirable sequences are encoded and optimizes uptake by cells when the construct encoding specific influenza virus genes is introduced into surrounding tissue. The specific DNA vectors of the present invention include but are not limited to V1, V1J (SEQ ID NO:13), V1Jneo (SEQ ID NO:14), V1Jns (Figure 1A, SEQ ID NO:15), V1R (SEQ ID NO:26), and any of the aforementioned vectors wherein a nucleotide sequence encoding a leader peptide. preferably the human tPA leader, is fused directly downstream of the CMV-intA promoter, including but not limited to V1Jns-tpa, as shown in Figure 1B and SEQ ID NO:28.

10

15

20

25

30

The present invention especially relates to a DNA vaccine and a pharmaceutically active vaccine composition which contains this DNA vaccine, and the use as prophylactic and/or therapeutic vaccine for host immunization, preferably human host immunization, against an HIV infection or to combat an existing HIV condition. These DNA vaccines are represented by codon optimized DNA molecules encoding HIV-1 Pol or biologically active Pol modifications or Pol-containing fusion proteins which are ligated within an appropriate DNA plasmid vector, with or without a nucleotide sequence encoding a functional leader peptide. DNA vaccines of the present invention may comprise codon optimized DNA molecules encoding HIV-1 Pol or biologically active Pol modifications or Pol-containing fusion proteins ligated in DNA vectors V1, V1J (SEQ ID NO:14), V1Jneo (SEQ ID NO:15), V1Jns (Figure 1A, SEQ ID NO:16), V1R (SEQ ID NO:26), or any of the aforementioned vectors

wherein a nucleotide sequence encoding a leader peptide, preferably the human tPA leader, is fused directly downstream of the CMV-intA promoter, including but not limited to V1Jns-tpa, as shown in Figure 1B and SEQ ID NO:28. To this end, polynucleotide vaccine constructions include, V1Jns-wtpol and V1R-wtpol (comprising the DNA molecule encoding WT Pol, as set forth in SEQ ID NO:2), V1Jns-tPA-WTPol, (comprising the DNA molecule encoding tPA Pol, as set forth in SEQ ID NO:6), V1Jns-IAPol (comprising the DNA molecule encoding IA Pol, as set forth in SEQ ID NO:4), and V1Jns-tPA-IAPol, (comprising the DNA molecule encoding tPA-IA Pol, as set forth in SEQ ID NO:8). Polynucleotide vaccine constructions V1R-wtpol, V1Jns-IAPol, and V1Jns-tPA-IAPol, are exemplified in Example Sections 3-5.

10

15

20

25

30

It will be evident upon review of the teaching within this specification that numerous vector/Pol antigen constructs may be generated. While the exemplified constructs are preferred, any number of vector/Pol antigen combinations are within the scope of the present invention, especially wild type or modified/inactivated Pol proteins which comprise at least one, preferably 5 or more and especially all nine mutations as shown in Table 1, with or without the inclusion of a leader sequence such as human tPA.

The DNA vector vaccines of the present invention may be formulated in any pharmaceutically effective formulation for host administration. Any such formulation may be, for example, a saline solution such as phosphate buffered saline (PBS). It will be useful to utilize pharmaceutically acceptable formulations which also provide long-term stability of the DNA vector vaccines of the present invention. During storage as a pharmaceutical entity, DNA plasmid vaccines undergo a physiochemical change in which the supercoiled plasmid converts to the open circular and linear form. A variety of storage conditions (low pH, high temperature, low ionic strength) can accelerate this process. Therefore, the removal and/or chelation of trace metal ions (with succinic or malic acid, or with chelators containing multiple phosphate ligands) from the DNA plasmid solution, from the formulation buffers or from the vials and closures, stabilizes the DNA plasmid from this degradation pathway during storage. In addition, inclusion of non-reducing free radical scavengers, such as ethanol or glycerol, are useful to prevent damage of the DNA plasmid from free radical production that may still occur, even in apparently demetalated solutions. Furthermore, the buffer type, pH, salt concentration, light

5

10

20 -

25

30

exposure, as well as the type of sterilization process used to prepare the vials, may be controlled in the formulation to optimize the stability of the DNA vaccine. Therefore, formulations that will provide the highest stability of the DNA vaccine will be one that includes a demetalated solution containing a buffer (phosphate or bicarbonate) with a pH in the range of 7-8, a salt (NaCl, KCl or LiCl) in the range of 100-200 mM, a metal ion chelator (e.g., EDTA, diethylenetriaminepenta-acetic acid (DTPA), malate, inositol hexaphosphate, tripolyphosphate or polyphosphoric acid), a nonreducing free radical scavenger (e.g. ethanol, glycerol, methionine or dimethyl sulfoxide) and the highest appropriate DNA concentration in a sterile glass vial, packaged to protect the highly purified, nuclease free DNA from light. A particularly preferred formulation which will enhance long term stability of the DNA vector vaccines of the present invention would comprise a Tris-HCl buffer at a pH from about 8.0 to about 9.0; ethanol or glycerol at about 3% w/v; EDTA or DTPA in a concentration range up to about 5 mM; and NaCl at a concentration from about 50 mM to about 500 mM. The use of such stabilized DNA vector vaccines and various alternatives to this preferred formulation range is described in detail in PCT International Application No. PCT/US97/06655 and PCT International Publication No. WO 97/40839, both of which are hereby incorporated by reference.

The DNA vector vaccines of the present invention may also be formulated with an adjuvant or adjuvants which may increase immunogenicity of the DNA polynucleotide vaccines of the present invention. A number of these adjuvants are known in the art and are available for use in a DNA vaccine, including but not limited to particle bombardment using DNA-coated gold beads, co-administration of DNA vaccines with plasmid DNA expressing cytokines, chemokines, or costimulatory molecules, formulation of DNA with cationic lipids or with experimental adjuvants such as saponin, monophosphoryl lipid A or other compounds which increase immunogenicity of the DNA vaccine. Another adjuvant for use in the DNA vector vaccines of the present invention are one or more forms of an aluminum phosphate-based adjuvant wherein the aluminum phosphate-based adjuvant possesses a molar PO<sub>4</sub>/Al ratio of approximately 0.9. An additional mineral-based adjuvant may be generated from one or more forms of a calcium phosphate. These mineral-based adjuvants are useful in increasing cellular and humoral responses to DNA vaccination. These mineral-based compounds for use as DNA vaccines adjuvants are disclosed in PCT International

Application No. PCT/US98/02414, PCT International Publication No. WO 98/35562, which is hereby incorporated by reference. Another preferred adjuvant is a non-ionic block copolymer which shows adjuvant activity with DNA vaccines. The basic structure comprises blocks of polyoxyethylene (POE) and polyoxypropylene (POP) such as a POE-POP-POE block copolymer. Newman et al. (1998, Critical Reviews in Therapeutic Drug Carrier Systems 15(2): 89-142) review a class of non-ionic block copolymers which show adjuvant activity. The basic structure comprises blocks of polyoxyethylene (POE) and polyoxypropylene (POP) such as a POE-POP-POE block copolymer. Newman et al. id., disclose that certain POE-POP-POE block copolymers may be useful as adjuvants to an influenza protein-based vaccine, namely higher molecular weight POE-POP-POE block copolymers containing a central POP block having a molecular weight of over about 9000 daltons to about 20,000 daltons and flanking POE blocks which comprise up to about 20% of the total molecular weight of the copolymer (see also U.S. Reissue Patent No. 36,665, U.S. Patent No. 5,567,859, U.S. Patent No. 5,691,387, U.S. Patent No. 5,696,298 and U.S. Patent No. 5,990,241, all issued to Emanuele, et al., regarding these POE-POP-POE block copolymers). WO 96/04932 further discloses higher molecular weight POE/POP block copolymers which have surfactant characteristics and show biological efficacy as vaccine adjuvants. The above cited references within this paragraph are hereby incorporated by reference in their entirety. It is therefore within the purview of the skilled artisan to utilize available adjuvants which may increase the immune response of the polynucleotide vaccines of the present invention in comparison to administration of a non-adjuvanted polynucleotide vaccine.

10

15

20

25

30

The DNA vector vaccines of the present invention are administered to the host by any means known in the art, such as enteral and parenteral routes. These routes of delivery include but are not limited to intramusclar injection, intraperitoneal injection, intravenous injection, inhalation or intranasal delivery, oral delivery, sublingual administration, subcutaneous administration, transdermal administration, transcutaneous administration, percutaneous administration or any form of particle bombardment, such as a biolostic device such as a "gene gun" or by any available needle-free injection device. The preferred methods of delivery of the HIV-1 Polbased DNA vaccines disclosed herein are intramuscular injection, subcutaneous administration and needle-free injection. An especially preferred method is

intramuscular delivery.

10

15

20

25

30

The amount of expressible DNA to be introduced to a vaccine recipient will depend on the strength of the transcriptional and translational promoters used in the DNA construct, and on the immunogenicity of the expressed gene product. In general, an immunologically or prophylactically effective dose of about 1  $\mu g$  to greater than about 20 mg, and preferably in doses from about 1 mg to about 5 mg is administered directly into muscle tissue. As noted above, subcutaneous injection, intradermal introduction, impression through the skin, and other modes of administration such as intraperitoneal, intravenous, inhalation and oral delivery are also contemplated. It is also contemplated that booster vaccinations are to be provided in a fashion which optimizes the overall immune response to the Pol-based DNA vector vaccines of the present invention.

The aforementioned polynucleotides, when directly introduced into a vertebrate *in vivo*, express the respective HIV-1 Pol protein within the animal and in turn induce a cellular immune response within the host to the expressed Pol antigen. To this end, the present invention also relates to methods of using the HIV-1 Polbased polynucleotide vaccines of the present invention to provide effective immunoprophylaxis, to prevent establishment of an HIV-1 infection following exposure to this virus, or as a post-HIV infection therapeutic vaccine to mitigate the acute HIV-1 infection so as to result in the establishment of a lower virus load with beneficial long term consequences. As noted above, the present invention contemplates a method of administration or use of the DNA pol-based vaccines of the present invention using an any of the known routes of introducing polynucleotides into living tissue to induce expression of proteins.

Therefore, the present invention provides for methods of using a DNA polbased vaccine utilizing the various parameters disclosed herein as well as any additional parameters known in the art, which, upon introduction into mammalian tissue induces intracellular expression of these DNA pol-based vaccines. This intracellular expression of the Pol-based immunogen induces a cellular immune response which provides a substantial level of protection against an existing HIV-1 infection or provides a substantial level of protection against a future infection in a presently uninfected host.

The following examples are provided to illustrate the present invention without, however, limiting the same hereto.

# **EXAMPLE 1**

# Vaccine Vectors

VI - Vaccine vector V1 was constructed from pCMVIE-AKI-DHFR (Whang et al., 1987, J. Virol. 61: 1796). The AKI and DHFR genes were removed by cutting 5 the vector with EcoRI and self-ligating. This vector does not contain intron A in the CMV promoter, so it was added as a PCR fragment that had a deleted internal SacI site [at 1855 as numbered in Chapman, et al., 1991, Nuc. Acids Res. 19: 3979). The template used for the PCR reactions was pCMVintA-Lux, made by ligating the HindIII and NheI fragment from pCMV6a120 (see Chapman et al., ibid.), which includes hCMV-IE1 enhancer/promoter and intron A, into the HindIII and XbaI sites of pBL3 to generate pCMVIntBL. The 1881 base pair luciferase gene fragment (HindIII-Smal Klenow filled-in) from RSV-Lux (de Wet et al., 1987, Mol. Cell Biol. 7: 725) was ligated into the Sall site of pCMVIntBL, which was Klenow filled-in and phosphatase treated. The primers that spanned intron A are: 5' primer: 5'-CTATAT 15 AAGCAGAGCTCGTTTAG-3' (SEQ ID NO:10); 3' primer: 5'-GTAGCAAA GATCTAAGGACGGTGACTGCAG-3' (SEQ ID NO:11). The primers used to remove the SacI site are: sense primer, 5'-GTATGTGTCTGAAAATGAGCG TGGAGATTGGGCTCGCAC-3' (SEQ ID NO:12) and the antisense primer, 5'-GTGCGAGCCCAATCTCCACGCTCATTTTCAGAC ACATAC-3' (SEO ID NO:13). The PCR fragment was cut with Sac I and Bgl II and inserted into the vector which had been cut with the same enzymes.

VIJ – Vaccine vector V1J was generated to remove the promoter and transcription termination elements from vector V1 in order to place them within a more defined context, create a more compact vector, and to improve plasmid purification yields. V1J is derived from vectors V1 and pUC18, a commercially available plasmid. V1 was digested with SspI and EcoRI restriction enzymes producing two fragments of DNA. The smaller of these fragments, containing the CMVintA promoter and Bovine Growth Hormone (BGH) transcription termination elements which control the expression of heterologous genes, was purified from an agarose electrophoresis gel. The ends of this DNA fragment were then "blunted" using the T4 DNA polymerase enzyme in order to facilitate its ligation to another "blunt-ended" DNA fragment. pUC18 was chosen to provide the "backbone" of the expression vector. It is known to produce high yields of plasmid, is well-

25

characterized by sequence and function, and is of small size. The entire lac operon was removed from this vector by partial digestion with the HaeII restriction enzyme. The remaining plasmid was purified from an agarose electrophoresis gel, blunt-ended with the T4 DNA polymerase treated with calf intestinal alkaline phosphatase, and ligated to the CMVintA/BGH element described above. Plasmids exhibiting either of two possible orientations of the promoter elements within the pUC backbone were obtained. One of these plasmids gave much higher yields of DNA in E. coli and was designated V1J. This vector's structure was verified by sequence analysis of the junction regions and was subsequently demonstrated to give comparable or higher 10 expression of heterologous genes compared with V1. The nucleotide sequence of V1J is as follows: TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG TCAGGGGCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC 15 ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG CTATTGGCCA TTGCATACGT TGTATCCATA TCATAATATG TACATTTATA TTGGCTCATG TCCAACATTA CCGCCATGTT GACATTGATT ATTGACTAGT TATTAATAGT AATCAATTAC GGGGTCATTA GTTCATAGCC CATATATGGA GTTCCGCGTT ACATAACTTA CGGTAAATGG CCCGCCTGGC TGACCGCCCA ACGACCCCCG CCCATTGACG TCAATAATGA CGTATGTTCC 20 CATAGTAACG CCAATAGGGA CTTTCCATTG ACGTCAATGG GTGGAGTATT TACGGTAAAC TGCCCACTTG GCAGTACATC AAGTGTATCA TATGCCAAGT ACGCCCCCTA TTGACGTCAA TGACGGTAAA TGGCCCGCCT GGCATTATGC CCAGTACATG ACCTTATGGG ACTTTCCTAC TTGGCAGTAC ATCTACGTAT TAGTCATCGC TATTACCATG GTGATGCGGT TTTGGCAGTA CATCAATGGG CGTGGATAGC GGTTTGACTC ACGGGGATTT CCAAGTCTCC ACCCCATTGA 25 CGTCAATGGG AGTTTGTTTT GGCACCAAAA TCAACGGGAC TTTCCAAAAT GTCGTAACAA CTCCGCCCA TTGACGCAAA TGGGCGGTAG GCGTGTACGG TGGGAGGTCT ATATAAGCAG AGCTCGTTTA GTGAACCGTC AGATCGCCTG GAGACGCCAT CCACGCTGTT TTGACCTCCA TAGAAGACAC CGGGACCGAT CCAGCCTCCG CGGCCGGGAA CGGTGCATTG GAACGCGGAT TCCCCGTGCC AAGAGTGACG TAAGTACCGC CTATAGAGTC TATAGGCCCA CCCCTTGGC 30 TTCTTATGCA TGCTATACTG TTTTTGGCTT GGGGTCTATA CACCCCCGCT TCCTCATGTT ATAGGTGATG GTATAGCTTA GCCTATAGGT GTGGGTTATT GACCATTATT GACCACTCCC CTATTGGTGA CGATACTTC CATTACTAAT CCATAACATG GCTCTTTGCC ACAACTCTCT TTATTGGCTA TATGCCAATA CACTGTCCTT CAGAGACTGA CACGGACTCT GTATTTTTAC

AGGATGGGGT CTCATTTATT ATTTACAAAT TCACATATAC AACACCACCG TCCCCAGTGC

	CCGCAGTTTT	TATTAAACAT	AACGTGGGAT	CTCCACGCGA	ATCTCGGGTA	CGTGTTCCGG
	ACATGGGCTC	TTCTCCGGTA	GCGGCGGAGC	TTCTACATCC	GAGCCCTGCT	CCCATGCCTC
	CAGCGACTCA	TGGTCGCTCG	GCAGCTCCTT	GCTCCTAACA	GTGGAGGCCA	GACTTAGGCA
	CAGCACGATG	CCCACCACCA	CCAGTGTGCC	GCACAAGGCC	GTGGCGGTAG	GGTATGTGTC
5	TGAAAATGAG	CTCGGGGAGC	GGGCTTGCAC	CGCTGACGCA	TTTGGAAGAC	TTAAGGCAGC
	GGCAGAAGAA	GATGCAGGCA	GCTGAGTTGT	TGTGTTCTGA	TAAGAGTCAG	AGGTAACTCC
	CGTTGCGGTG	CTGTTAACGG	TGGAGGGCAG	TGTAGTCTGA	GCAGTACTCG	TTGCTGCCGC
	GCGCGCCACC	AGACATAATA	GCTGACAGAC	TAACAGACTG	TTCCTTTCCA	TGGGTCTTTT
	CTGCAGTCAC	CGTCCTTAGA	TCTGCTGTGC	CTTCTAGTTG	CCAGCCATCT	GTTGTTTGCC
10	CCTCCCCGT	GCCTTCCTTG	ACCCTGGAAG	GTGCCACTCC	CACTGTCCTT	TCCTAATAAA
	ATGAGGAAAT	TGCATCGCAT	TGTCTGAGTA	GGTGTCATTC	TATTCTGGGG	GGTGGGGTGG
	GGCAGCACAG	CAAGGGGGAG	GATTGGGAAG	ACAATAGCAG	GCATGCTGGG	GATGCGGTGG
	GCTCTATGGG	TACCCAGGTG	CTGAAGAATT	GACCCGGTTC	CTCCTGGGCC	AGAAAGAAGC
	AGGCACATCC	CCTTCTCTGT	GACACACCCT	GTCCACGCCC	CTGGTTCTTA	GTTCCAGCCC
15	CACTCATAGG	ACACTCATAG	CTCAGGAGGG	CTCCGCCTTC	AATCCCACCC	GCTAAAGTAC
	TTGGAGCGGT	CTCTCCCTCC	CTCATCAGCC	CACCAAACCA	AACCTAGCCT	CCAAGAGTGG
	GAAGAAATTA	AAGCAAGATA	GGCTATTAAG	TGCAGAGGGA	GAGAAAATGC	CTCCAACATG
	TGAGGAAGTA	ATGAGAGAAA	TCATAGAATT	TCTTCCGCTT	CCTCGCTCAC	TGACTCGCTG
	CGCTCGGTCG	TTCGGCTGCG	GCGAGCGGTA	TCAGCTCACT	CAAAGGCGGT	AATACGGTTA
20	TCCACAGAAT	CAGGGGATAA	CGCAGGAAAG	AACATGTGAG	CAAAAGGCCA	GCAAAAGGCC
	AGGAACCGTA	AAAAGGCCGC	GTTGCTGGCG	TTTTTCCATA	GGCTCCGCCC	CCCTGACGAG
	CATCACAAAA	ATCGACGCTC	AAGTCAGAGG	TGGCGAAACC	CGACAGGACT	ATAAAGATAC
	CAGGCGTTTC	CCCCTGGAAG	CTCCCTCGTG	CGCTCTCCTG	TTCCGACCCT	GCCGCTTACC
	GGATACCTGT	CCGCCTTTCT	CCCTTCGGGA	AGCGTGGCGC	TTTCTCAATG	CTCACGCTGT
25	AGGTATCTCA	GTTCGGTGTA	GGTCGTTCGC	TCCAAGCTGG	GCTGTGTGCA	CGAACCCCCC
	GTTCAGCCCG	ACCGCTGCGC	CTTATCCGGT	AACTATCGTC	TTGAGTCCAA	CCCGGTAAGA
	CACGACTTAT	CGCCACTGGC	AGCAGCCACT	GGTAACAGGA	TTAGCAGAGC	GAGGTATGTA
	GGCGGTGCTA	CAGAGTTCTT	GAAGTGGTGG	CCTAACTACG	GCTACACTAG	AAGGACAGTA
	TTTGGTATCT	GCGCTCTGCT	GAAGCCAGTT	ACCTTCGGAA	AAAGAGTTGG	TAGCTCTTGA
30	TCCGGCAAAC	AAACCACCGC	TGGTAGCGGT	GGTTTTTTTG	TTTGCAAGCA	GCAGATTACG
	CGCAGAAAAA	AAGGATCTCA	AGAAGATCCT	TTGATCTTTT	CTACGGGGTC	TGACGCTCAG
	TGGAACGAAA	ACTCACGTTA	AGGGATTTTG	GTCATGAGAT	TATCAAAAAG	GATCTTCACC
	TAGATCCTTT	TAAATTAAAA	ATGAAGTTTT	AAATCAATCT	AAAGTATATA	TGAGTAAACT
	TGGTCTGACA	GTTACCAATG	CTTAATCAGT	GAGGCACCTA	TCTCAGCGAT	CTGTCTATTT

CGTTCATCCA TAGTTGCCTG ACTCCCCGTC GTGTAGATAA CTACGATACG GGAGGGCTTA CCATCTGGCC CCAGTGCTGC AATGATACCG CGAGACCCAC GCTCACCGGC TCCAGATTTA TCAGCAATAA ACCAGCCAGC CGGAAGGGCC GAGCGCAGAA GTGGTCCTGC AACTTTATCC GCCTCCATCC AGTCTATTAA TTGTTGCCGG GAAGCTAGAG TAAGTAGTTC GCCAGTTAAT AGTTTGCGCA ACGTTGTTGC CATTGCTACA GGCATCGTGG TGTCACGCTC GTCGTTTGGT ATGGCTTCAT TCAGCTCCGG TTCCCAACGA TCAAGGCGAG TTACATGATC CCCCATGTTG TGCAAAAAG CGGTTAGCTC CTTCGGTCCT CCGATCGTTG TCAGAAGTAA GTTGGCCGCA GTGTTATCAC TCATGGTTAT GGCAGCACTG CATAATTCTC TTACTGTCAT GCCATCCGTA AGATGCTTTT CTGTGACTGG TGAGTACTCA ACCAAGTCAT TCTGAGAATA GTGTATGCGG CGACCGAGTT GCTCTTGCCC GGCGTCAATA CGGGATAATA CCGCGCCACA TAGCAGAACT TTAAAAGTGC TCATCATTGG AAAACGTTCT TCGGGGCGAA AACTCTCAAG GATCTTACCG CTGTTGAGAT CCAGTTCGAT GTAACCCACT CGTGCACCCA ACTGATCTTC AGCATCTTTT ACTITCACCA GCGITTCTGG GTGAGCAAAA ACAGGAAGGC AAAATGCCGC AAAAAAGGGA ATAAGGGCGA CACGGAAATG TTGAATACTC ATACTCTTCC TTTTTCAATA TTATTGAAGC ATTTATCAGG GTTATTGTCT CATGAGCGGA TACATATTTG AATGTATTTA GAAAAATAAA CAAATAGGGG TTCCGCGCAC ATTTCCCCGA AAAGTGCCAC CTGACGTCTA AGAAACCATT ATTATCATGA CATTAACCTA TAAAAATAGG CGTATCACGA GGCCCTTTCG TC (SEQ ID NO:14).

10

15

VIJneo - Construction of vaccine vector VIJneo expression vector involved removal of the ampr gene and insertion of the kanr gene (neomycin 20 phosphotransferase). The amp<sup>r</sup> gene from the pUC backbone of V1J was removed by digestion with SspI and Eam1105I restriction enzymes. The remaining plasmid was purified by agarose gel electrophoresis, blunt-ended with T4 DNA polymerase, and then treated with calf intestinal alkaline phosphatase. The commercially available kan<sup>r</sup> gene, derived from transposon 903 and contained within the pUC4K plasmid, 25 was excised using the PstI restriction enzyme, purified by agarose gel electrophoresis, and blunt-ended with T4 DNA polymerase. This fragment was ligated with the V1J backbone and plasmids with the kan<sup>r</sup> gene in either orientation were derived which were designated as V1Jneo #'s 1 and 3. Each of these plasmids was confirmed by restriction enzyme digestion analysis, DNA sequencing of the junction regions, and was shown to produce similar quantities of plasmid as V1J. Expression of heterologous gene products was also comparable to V1J for these V1Jneo vectors. V1Ineo#3, referred to as V1Ineo hereafter, was selected which contains the kan<sup>r</sup> gene in the same orientation as the ampr gene in V1J as the expression construct and

provides resistance to neomycin, kanamycin and G418. The nucleotide sequence of V1Jneo is as follows:

TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG TCAGGGGCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG CTATTGGCCA TTGCATACGT TGTATCCATA TCATAATATG TACATTTATA TTGGCTCATG TCCAACATTA CCGCCATGTT GACATTGATT ATTGACTAGT TATTAATAGT AATCAATTAC GGGGTCATTA GTTCATAGCC CATATATGGA GTTCCGCGTT ACATAACTTA CGGTAAATGG 10 CCCGCCTGGC TGACCGCCCA ACGACCCCCG CCCATTGACG TCAATAATGA CGTATGTTCC CATAGTAACG CCAATAGGGA CTTTCCATTG ACGTCAATGG GTGGAGTATT TACGGTAAAC TGCCCACTTG GCAGTACATC AAGTGTATCA TATGCCAAGT ACGCCCCCTA TTGACGTCAA TGACGGTAAA TGGCCCGCCT GGCATTATGC CCAGTACATG ACCTTATGGG ACTTTCCTAC TTGGCAGTAC ATCTACGTAT TAGTCATCGC TATTACCATG GTGATGCGGT TTTGGCAGTA CATCAATGGG CGTGGATAGC GGTTTGACTC ACGGGGATTT CCAAGTCTCC ACCCCATTGA CGTCAATGGG AGTTTGTTTT GGCACCAAAA TCAACGGGAC TTTCCAAAAT GTCGTAACAA CTCCGCCCCA TTGACGCAAA TGGGCGGTAG GCGTGTACGG TGGGAGGTCT ATATAAGCAG AGCTCGTTTA GTGAACCGTC AGATCGCCTG GAGACGCCAT CCACGCTGTT TTGACCTCCA TAGAAGACAC CGGGACCGAT CCAGCCTCCG CGGCCGGGAA CGGTGCATTG GAACGCGGAT 20 TCCCCGTGCC AAGAGTGACG TAAGTACCGC CTATAGAGTC TATAGGCCCA CCCCCTTGGC TTCTTATGCA TGCTATACTG TTTTTGGCTT GGGGTCTATA CACCCCGGCT TCCTCATGTT ATAGGTGATG GTATAGCTTA GCCTATAGGT GTGGGTTATT GACCATTATT GACCACTCCC CTATTGGTGA CGATACTTTC CATTACTAAT CCATAACATG GCTCTTTGCC ACAACTCTCT TTATTGGCTA TATGCCAATA CACTGTCCTT CAGAGACTGA CACGGACTCT GTATTTTTAC 25 AGGATGGGGT CTCATTTATT ATTTACAAAT TCACATATAC AACACCACCG TCCCCAGTGC CCGCAGTTTT TATTAAACAT AACGTGGGAT CTCCACGCGA ATCTCGGGTA CGTGTTCCGG ACATGGGCTC TTCTCCGGTA GCGGCGGAGC TTCTACATCC GAGCCCTGCT CCCATGCCTC CAGCGACTCA TGGTCGCTCG GCAGCTCCTT GCTCCTAACA GTGGAGGCCA GACTTAGGCA CAGCACGATG CCCACCACCA CCAGTGTGCC GCACAAGGCC GTGGCGGTAG GGTATGTGTC TGAAAATGAG CTCGGGGAGC GGGCTTGCAC CGCTGACGCA TTTGGAAGAC TTAAGGCAGC GGCAGAAGAA GATGCAGGCA GCTGAGTTGT TGTGTTCTGA TAAGAGTCAG AGGTAACTCC CGTTGCGGTG CTGTTAACGG TGGAGGGCAG TGTAGTCTGA GCAGTACTCG TTGCTGCCGC GCGCGCCACC AGACATAATA GCTGACAGAC TAACAGACTG TTCCTTTCCA TGGGTCTTTT CTGCAGTCAC CGTCCTTAGA TCTGCTGTGC CTTCTAGTTG CCAGCCATCT GTTGTTTGCC

	CCTCCCCGT	GCCTTCCTTG	ACCCTGGAAG	GTGCCACTCC	CACTGTCCTT	TCCTAATAAA
	ATGAGGAAAT	TGCATCGCAT	TGTCTGAGTA	GGTGTCATTC	TATTCTGGGG	GGTGGGGTGG
	GGCAGCACAG	CAAGGGGGAG	GATTGGGAAG	ACAATAGCAG	GCATGCTGGG	GATGCGGTGG
	GCTCTATGGG	TACCCAGGTG	CTGAAGAATT	GACCCGGTTC	CTCCTGGGCC	AGAAAGAAGC
5	AGGCACATCC	CCTTCTCTGT	GACACACCCT	GTCCACGCCC	CTGGTTCTTA	GTTCCAGCCC
	CACTCATAGG	ACACTCATAG	CTCAGGAGGG	CTCCGCCTTC	AATCCCACCC	GCTAAAGTAC
	TTGGAGCGGT	CTCTCCCTCC	CTCATCAGCC	CACCAAACCA	AACCTAGCCT	CCAAGAGTGG
	GAAGAAATTA	AAGCAAGATA	GGCTATTAAG	TGCAGAGGGA	GAGAAAATGC	CTCCAACATG
	TGAGGAAGTA	ATGAGAGAAA	TCATAGAATT	TCTTCCGCTT	CCTCGCTCAC	TGACTCGCTG
10	CGCTCGGTCG	TTCGGCTGCG	GCGAGCGGTA	TCAGCTCACT	CAAAGGCGGT	AATACGGTTA
	TCCACAGAAT	CAGGGGATAA	CGCAGGAAAG	AACATGTGAG	CAAAAGGCCA	GCAAAAGGCC
	AGGAACCGTA	AAAAGGCCGC	GTTGCTGGCG	TTTTTCCATA	GGCTCCGCCC	CCCTGACGAG
	CATCACAAAA	ATCGACGCTC	AAGTCAGAGG	TGGCGAAACC	CGACAGGACT	ATAAAGATAC
	CAGGCGTTTC	CCCCTGGAAG	CTCCCTCGTG	CGCTCTCCTG	TTCCGACCCT	GCCGCTTACC
15	GGATACCTGT	CCGCCTTTCT	CCCTTCGGGA	AGCGTGGCGC	TTTCTCAATG	CTCACGCTGT
	AGGTATCTCA	GTTCGGTGTA	GGTCGTTCGC	TCCAAGCTGG	GCTGTGTGCA	CGAACCCCCC
	GTTCAGCCCG	ACCGCTGCGC	CTTATCCGGT	AACTATCGTC	TTGAGTCCAA	CCCGGTAAGA
	CACGACTTAT	CGCCACTGGC	AGCAGCCACT	GGTAACAGGA	TTAGCAGAGC	GAGGTATGTA
	GGCGGTGCTA	CAGAGTTCTT	GAAGTGGTGG	CCTAACTACG	GCTACACTAG	AAGGACAGTA
20	TTTGGTATCT	GCGCTCTGCT	GAAGCCAGTT	ACCTTCGGAA	AAAGAGTTGG	TAGCTCTTGA
	TCCGGCAAAC	AAACCACCGC	TGGTAGCGGT	GGTTTTTTTG	TTTGCAAGCA	GCAGATTACG
	CGCAGAAAAA	AAGGATCTCA	AGAAGATCCT	TTGATCTTTT	CTACGGGGTC	TGACGCTCAG
	TGGAACGAAA	ACTCACGTTA	AGGGATTTTG	GTCATGAGAT	TATCAAAAAG	GATCTTCACC
	TAGATCCTTT	TAAATTAAAA	ATGAAGTTTT	AAATCAATCT	AAAGTATATA	TGAGTAAACT
25	TGGTCTGACA	GTTACCAATG	CTTAATCAGT	GAGGCACCTA	TCTCAGCGAT	CTGTCTATTT
	CGTTCATCCA	TAGTTGCCTG	ACTCCGGGGG	GGGGGGGCGC	TGAGGTCTGC	CTCGTGAAGA
	AGGTGTTGCT	GACTCATACC	AGGCCTGAAT	CGCCCCATCA	TCCAGCCAGA	AAGTGAGGGA
	GCCACGGTTG	ATGAGAGCTT	TGTTGTAGGT	GGACCAGTTG	GTGATTTTGA	ACTTTTGCTT
	TGCCACGGAA	CGGTCTGCGT	TGTCGGGAAG	ATGCGTGATC	TGATCCTTCA	ACTCAGCAAA
30	AGTTCGATTT	ATTCAACAAA	GCCGCCGTCC	CGTCAAGTCA	GCGTAATGCT	CTGCCAGTGT
	TACAACCAAT	TAACCAATTC	TGATTAGAAA	AACTCATCGA	GCATCAAATG	AAACTGCAAT
	TTATTCATAT	CAGGATTATC	AATACCATAT	TTTTGAAAAA	GCCGTTTCTG	TAATGAAGGA
	GAAAACTCAC	CGAGGCAGTT	CCATAGGATG	GCAAGATCCT	GGTATCGGTC	TGCGATTCCG
	ACTCGTCCAA	CATCAATACA	ACCTATTAAT	TTCCCCTCGT	CAAAAATAAG	GTTATCAAGT

GAGAAATCAC CATGAGTGAC GACTGAATCC GGTGAGAATG GCAAAAGCTT ATGCATTTCT TTCCAGACTT GTTCAACAGG CCAGCCATTA CGCTCGTCAT CAAAATCACT CGCATCAACC AAACCGTTAT TCATTCGTGA TTGCGCCTGA GCGAGACGAA ATACGCGATC GCTGTTAAAA GGACAATTAC AAACAGGAAT CGAATGCAAC CGGCGCAGGA ACACTGCCAG CGCATCAACA ATATTTCAC CTGAATCAGG ATATTCTTCT AATACCTGGA ATGCTGTTTT CCCGGGGATC GCAGTGGTGA GTAACCATGC ATCATCAGGA GTACGGATAA AATGCTTGAT GGTCGGAAGA GGCATAAATT CCGTCAGCCA GTTTAGTCTG ACCATCTCAT CTGTAACATC ATTGGCAACG CTACCTTTGC CATGTTTCAG AAACAACTCT GGCGCATCGG GCTTCCCATA CAATCGATAG ATTGTCGCAC CTGATTGCCC GACATTATCG CGAGCCCATT TATACCCATA TAAATCAGCA TCCATGTTGG AATTTAATCG CGGCCTCGAG CAAGACGTTT CCCGTTGAAT ATGGCTCATA ACACCCCTTG TATTACTGTT TATGTAAGCA GACAGTTTTA TTGTTCATGA TGATATATTT TTATCTTGTG CAATGTAACA TCAGAGATTT TGAGACACAA CGTGGCTTTC CCCCCCCCC CATTATTGAA GCATTTATCA GGGTTATTGT CTCATGAGCG GATACATATT TGAATGTATT TAGAAAATA AACAAATAGG GGTTCCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC GAGGCCCTTT CGTC (SEQ ID NO:15).

V1Jns - The expression vector VIJns was generated by adding an SfiI site to V1Jneo to facilitate integration studies. A commercially available 13 base pair SfiI linker (New England BioLabs) was added at the KpnI site within the BGH sequence of the vector. V1Jneo was linearized with KpnI, gel purified, blunted by T4 DNA polymerase, and ligated to the blunt SfiI linker. Clonal isolates were chosen by restriction mapping and verified by sequencing through the linker. The new vector was designated V1Jns. Expression of heterologous genes in V1Jns (with SfiI) was comparable to expression of the same genes in V1Jneo (with KpnI).

The nucleotide sequence of V1Jns is as follows:

5

10

15

20

30

TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA
CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG TCAGGGCGCG TCAGCGGGTG
TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC
ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG
CTATTGGCCA TTGCATACGT TGTATCCATA TCATAATATG TACATTTATA TTGGCTCATG
TCCAACATTA CCGCCATGTT GACATTGATT ATTGACTAGT TATTAATAGT AATCAATTAC
GGGGTCATTA GTTCATAGCC CATATATGGA GTTCCGCGTT ACATAACTTA CGGTAAATGG
CCCGCCTGGC TGACCGCCCA ACGACCCCCG CCCATTGACG TCAATAATGA CGTATATCC
CATAGTAACG CCCAATAGGGA CTTTCCATTG ACGTCAATGG GTGGAGTATT TACGGTAAAC

	TGCCCACTTG	GCAGTACATC	AAGTGTATCA	TATGCCAAGT	ACGCCCCCTA	TTGACGTCAA
	TGACGGTAAA	TGGCCCGCCT	GGCATTATGC	CCAGTACATG	ACCTTATGGG	ACTTTCCTAC
	TTGGCAGTAC	ATCTACGTAT	TAGTCATCGC	TATTACCATG	GTGATGCGGT	TTTGGCAGTA
	CATCAATGGG	CGTGGATAGC	GGTTTGACTC	ACGGGGATTT	CCAAGTCTCC	ACCCCATTGA
5	CGTCAATGGG	AGTTTGTTTT	GGCACCAAAA	TCAACGGGAC	TTTCCAAAAT	GTCGTAACAA
	CTCCGCCCCA	TTGACGCAAA	TGGGCGGTAG	GCGTGTACGG	TGGGAGGTCT	ATATAAGCAG
	AGCTCGTTTA	GTGAACCGTC	AGATCGCCTG	GAGACGCCAT	CCACGCTGTT	TTGACCTCCA
	TAGAAGACAC	CGGGACCGAT	CCAGCCTCCG	CGGCCGGGAA	CGGTGCATTG	GAACGCGGAT
	TCCCCGTGCC	AAGAGTGACG	TAAGTACCGC	CTATAGACTC	TATAGGCACA	CCCCTTTGGC
10	TCTTATGCAT	GCTATACTGT	TTTTGGCTTG	GGGCCTATAC	ACCCCCGCTT	CCTTATGCTA
	TAGGTGATGG	TATAGCTTAG	CCTATAGGTG	TGGGTTATTG	ACCATTATTG	ACCACTCCCC
	TATTGGTGAC	GATACTTTCC	ATTACTAATC	CATAACATGG	CTCTTTGCCA	CAACTATCTC
•	TATTGGCTAT	ATGCCAATAC	TCTGTCCTTC	AGAGACTGAC	ACGGACTCTG	TATTTTTACA
	GGATGGGGTC	CCATTTATTA	TTTACAAATT	CACATATACA	ACAACGCCGT	CCCCCGTGCC
15	CGCAGTTTTT	ATTAAACATA	GCGTGGGATC	TCCACGCGAA	TCTCGGGTAC	GTGTTCCGGA
	CATGGGCTCT	TCTCCGGTAG	CGGCGGAGCT	TCCACATCCG	AGCCCTGGTC	CCATGCCTCC
	AGCGGCTCAT	GGTCGCTCGG	CAGCTCCTTG	CTCCTAACAG	TGGAGGCCAG	ACTTAGGCAC
	AGCACAATGC	CCACCACCAC	CAGTGTGCCG	CACAAGGCCG	TGGCGGTAGG	GTATGTGTCT
	GAAAATGAGC	GTGGAGATTG	GGCTCGCACG	GCTGACGCAG	ATGGAAGACT	TAAGGCAGCG.
20	GCAGAAGAAG	ATGCAGGCAG	CTGAGTTGTT	GTATTCTGAT	AAGAGTCAGA	GGTAACTCCC
	GTTGCGGTGC	TGTTAACGGT	GGAGGGCAGT	GTAGTCTGAG	CAGTACTCGT	TGCTGCCGCG
	CGCGCCACCA	GACATAATAG	CTGACAGACT	AACAGACTGT	TCCTTTCCAT	GGGTCTTTTC
	TGCAGTCACC	GTCCTTAGAT	CTGCTGTGCC	TTCTAGTTGC	CAGCCATCTG	TTGTTTGCCC
	CTCCCCCGTG	CCTTCCTTGA	CCCTGGAAGG	TGCCACTCCC	ACTGTCCTTT	CCTAATAAAA
<b>25</b> .	TGAGGAAATT	GCATCGCATT	GTCTGAGTAG	GTGTCATTCT	ATTCTGGGGG	GTGGGGTGGG.
	GCAGGACAGC	AAGGGGGAGG	ATTGGGAAGA	CAATAGCAGG	CATGCTGGGG	ATGCGGTGGG
	CTCTATGGCC	GCTGCGGCCA	GGTGCTGAAG	AATTGACCCG	GTTCCTCCTG	GGCCAGAAAG
	AAGCAGGCAC	ATCCCCTTCT	CTGTGACACA	CCCTGTCCAC	GCCCCTGGTT	CTTAGTTCCA
	GCCCCACTCA	TAGGACACTC	ATAGCTCAGG	AGGGCTCCGC	CTTCAATCCC	ACCCGCTAAA
30	GTACTTGGAG	CGGTCTCTCC	CTCCCTCATC	AGCCCACCAA	ACCAAACCTA	GCCTCCAAGA
	GTGGGAAGAA	ATTAAAGCAA	GATAGGCTAT	TAAGTGCAGA	GGGAGAGAAA	ATGCCTCCAA
	CATGTGAGGA	AGTAATGAGA	GAAATCATAG	AATTTCTTCC	GCTTCCTCGC	TCACTGACTC
	GCTGCGCTCG	GTCGTTCGGC	TGCGGCGAGC	GGTATCAGCT	CACTCAAAGG	CGGTAATACG
	GTTATCCACA	GAATCAGGGG	ATAACGCAGG	AAAGAACATG	TGAGCAAAAG	GCCAGCAAAA

	GGCCAGGAAC	CGTAAAAAGG	CCGCGTTGCT	GGCGTTTTTC	CATAGGCTCC	GCCCCCTGA
	CGAGCATCAC	AAAAATCGAC	GCTCAAGTCA	GAGGTGGCGA	AACCCGACAG	GACTATAAAG
	ATACCAGGCG	TTTCCCCCTG	GAAGCTCCCT	CGTGCGCTCT	CCTGTTCCGA	CCCTGCCGCT
	TACCGGATAC	CTGTCCGCCT	TTCTCCCTTC	GGGAAGCGTG	GCGCTTTCTC	ATAGCTCACG
5	CTGTAGGTAT	CTCAGTTCGG	TGTAGGTCGT	TCGCTCCAAG	CTGGGCTGTG	TGCACGAACC
	CCCCGTTCAG	CCCGACCGCT	GCGCCTTATC	CGGTAACTAT	CGTCTTGAGT	CCAACCCGGT
	AAGACACGAC	TTATCGCCAC	TGGCAGCAGC	CACTGGTAAC	AGGATTAGCA	GAGCGAGGTA
	TGTAGGCGGT	GCTACAGAGT	TCTTGAAGTG	GTGGCCTAAC	TACGGCTACA	CTAGAAGAAC
	AGTATTTGGT	ATCTGCGCTC	TGCTGAAGCC	AGTTACCTTC	GGAAAAAGAG	TTGGTAGCTC
10	TTGATCCGGC	AAACAAACCA	CCGCTGGTAG	CGGTGGTTTT	TTTGTTTGCA	AGCAGCAGAT
	TACGCGCAGA	AAAAAAGGAT	CTCAAGAAGA	TCCTTTGATC	TTTTCTACGG	GGTCTGACGC
	TCAGTGGAAC	GAAAACTCAC	GTTAAGGGAT	TTTGGTCATG	AGATTATCAA	AAAGGATCTT
	CACCTAGATC	CTTTTAAATT	AAAAATGAAG	TTTTAAATCA	ATCTAAAGTA	TATATGAGTA
	AACTTGGTCT	GACAGTTACC	AATGCTTAAT	CAGTGAGGCA	CCTATCTCAG	CGATCTGTCT
15	ATTTCGTTCA	TCCATAGTTG	CCTGACTCGG	GGGGGGGGG	CGCTGAGGTC	TGCCTCGTGA
	AGAAGGTGTT	GCTGACTCAT	ACCAGGCCTG	AATCGCCCCA	TCATCCAGCC	AGAAAGTGAG
	GGAGCCACGG	TTGATGAGAG	CTTTGTTGTA	GGTGGACCAG	TTGGTGATTT	TGAACTTTTG ·
	CTTTGCCACG	GAACGGTCTG	CGTTGTCGGG	AAGATGCGTG	ATCTGATCCT	TCAACTCAGC
	AAAAGTTCGA	TTTATTCAAC	AAAGCCGCCG	TCCCGTCAAG	TCAGCGTAAT	GCTCTGCCAG
20	TGTTACAACC	AATTAACCAA	TTCTGATTAG	AAAAACTCAT	CGAGCATCAA	ATGAAACTGC
	AATTTATTCA	TATCAGGATT	ATCAATACCA	TATTTTTGAA	AAAGCCGTTT	CTGTAATGAA
	GGAGAAAACT	CACCGAGGCA	GTTCCATAGG	ATGGCAAGAT	CCTGGTATCG	GTCTGCGATT
	CCGACTCGTC	CAACATCAAT	ACAACCTATT	AATTTCCCCT	CGTCAAAAAT	AAGGTTATCA
	AGTGAGAAAT	CACCATGAGT	GACGACTGAA	TCCGGTGAGA	ATGGCAAAAG	CTTATGCATT
25	TCTTTCCAGA	CTTGTTCAAC	AGGCCAGCCA	TTACGCTCGT	CATCAAAATC	ACTCGCATCA
	ACCAAACCGT	TATTCATTCG	TGATTGCGCC	TGAGCGAGAC	GAAATACGCG	ATCGCTGTTA
	AAAGGACAAT	TACAAACAGG	AATCGAATGC	AACCGGCGCA	GGAACACTGC	CAGCGCATCA
	ACAATATTTT	CACCTGAATC	AGGATATTCT	TCTAATACCT	GGAATGCTGT	TTTCCCGGGG
	ATCGCAGTGG	TGAGTAACCA	TGCATCATCA	GGAGTACGGA	TAAAATGCTT	GATGGTCGGA
30	AGAGGCATAA	ATTCCGTCAG	CCAGTTTAGT	CTGACCATCT	CATCTGTAAC	ATCATTGGCA
	ACGCTACCTT	TGCCATGTTT	CAGAAACAAC	TCTGGCGCAT	CGGGCTTCCC	ATACAATCGA
	TAGATTGTCG	CACCTGATTG	CCCGACATTA	TCGCGAGCCC	ATTTATACCC	АТАТАААТСА
	GCATCCATGT	TGGAATTTAA	TCGCGGCCTC	GAGCAAGACG	TTTCCCGTTG	AATATGGCTC
	ATAACACCCC	TTGTATTACT	GTTTATGTAA	GCAGACAGTT	TTATTGTTCA	TGATGATATA

TTTTTATCTT GTGCAATGTA ACATCAGAGA TTTTGAGACA CAACGTGGCT TTCCCCCCCC CCCCATTATT GAAGCATTTA TCAGGGTTAT TGTCTCATGA GCGGATACAT ATTTGAATGT ATTTAGAAAA ATAAACAAAT AGGGGTTCCG CGCACATTTC CCCGAAAAGT GCCACCTGAC GTCTAAGAAA CCATTATTAT CATGACATTA ACCTATAAAA ATAGGCGTAT CACGAGGCCC TTTCGTC (SEQ ID NO:16).

The underlined nucleotides of SEQ ID NO:16 represent the Sfi1 site introduced into the Kpn 1 site of V1Ineo.

5

30

V1Jns-tPA - The vaccine vector V1Jns-tPA was constructed in order to fuse an heterologous leader peptide sequence to the pol DNA constructs of the present 10 invention. More specifically, the vaccine vector V1Jns was modified to include the human tissue-specific plasminogen activator (tPA) leader. As an exemplification, but by no means a limitation of generating a pol DNA construct comprising an aminoterminal leader sequence, plasmid V1Jneo was modified to include the human tissuespecific plasminogen activator (tPA) leader. Two synthetic complementary oligomers 15 were annealed and then ligated into V1Jneo which had been BglII digested. The sense and antisense oligomers were 5'-GATCACCATGGATGCAATGAAGAG AGGGCTCTGCTGTGCTGCTGCTGTGGAGCAGTCTTCGTTTCGCCCAG CGA-3' (SEQ ID NO:17); and, 5'-GATCTCGCTGGGCGAAACGAAGACTGCTCC ACACAGCAGCACACAGCAGAGCCCTCTCTTCATTGCATCCATGGT-3' 20 (SEQ ID NO:18). The Kozak sequence is underlined in the sense oligomer. These oligomers have overhanging bases compatible for ligation to BglII-cleaved sequences. After ligation the upstream BglII site is destroyed while the downstream BglII is retained for subsequent ligations. Both the junction sites as well as the entire tPA leader sequence were verified by DNA sequencing. Additionally, in order to conform 25 with V1Jns (=V1Jneo with an SfiI site), an SfiI restriction site was placed at the KpnI site within the BGH terminator region of V1Jneo-tPA by blunting the KpnI site with T4 DNA polymerase followed by ligation with an SfiI linker (catalogue #1138, New England Biolabs), resulting in V1Jns-tPA. This modification was verified by restriction digestion and agarose gel electrophoresis.

The V1Jns-tpa vector nucleotide sequence is as follows:

TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA

CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG TCAGGGCGCG TCAGCGGGTG

TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC

ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG

•	CTATTGGCCA	TTGCATACGT	TGTATCCATA	TCATAATATG	TACATTTATA	TTGGCTCATG
	TCCAACATTA	CCGCCATGTT	GACATTGATT	ATTGACTAGT	TATTAATAGT	AATCAATTAC
	GGGGTCATTA	GTTCATAGCC	CATATATGGA	GTTCCGCGTT	ACATAACTTA	CGGTAAATGG
	CCCGCCTGGC	TGACCGCCCA	ACGACCCCCG	CCCATTGACG	TCAATAATGA	CGTATGTTCC
5	CATAGTAACG	CCAATAGGGA	CTTTCCATTG	ACGTCAATGG	GTGGAGTATT	TACGGTAAAC
	TGCCCACTTG	GCAGTACATC	AAGTGTATCA	TATGCCAAGT	ACGCCCCTA	TTGACGTCAA
	TGACGGTAAA	TGGCCCGCCT	GGCATTATGC	CCAGTACATG	ACCTTATGGG	ACTTTCCTAC
	TTGGCAGTAC	ATCTACGTAT	TAGTCATCGC	TATTACCATG	GTGATGCGGT	TTTGGCAGTA
	CATCAATGGG	CGTGGATAGC	GGTTTGACTC	ACGGGGATTT	CCAAGTCTCC	ACCCCATTGA
10	CGTCAATGGG	AGTTTGTTTT	GGCACCAAAA	TCAACGGGAC	TTTCCAAAAT	GTCGTAACAA
	CTCCGCCCCA	TTGACGCAAA	TGGGCGGTAG	GCGTGTACGG	TGGGAGGTCT	ATATAAGCAG
	AGCTCGTTTA	GTGAACCGTC	AGATCGCCTG	GAGACGCCAT	CCACGCTGTT	TTGACCTCCA
•	TAGAAGACAC	CGGGACCGAT	CCAGCCTCCG	CGGCCGGGAA	CGGTGCATTG	GAACGCGGAT
	TCCCCGTGCC	AAGAGTGACG	TAAGTACCGC	CTATAGACTC	TATAGGCACA	CCCCTTTGGC
15	TCTTATGCAT	GCTATACTGT	TTTTGGCTTG	GGGCCTATAC	ACCCCCGCTT	CCTTATGCTA
	TAGGTGATGG	TATAGCTTAG	CCTATAGGTG	TGGGTTATTG	ACCATTATTG	ACCACTCCCC
	TATTGGTGAC	GATACTTTCC	ATTACTAATC	CATAACATGG	CTCTTTGCCA	CAACTATCTC
	TATTGGCTAT	ATGCCAATAC	TCTGTCCTTC	AGAGACTGAC	ACGGACTCTG	TATTTTTACA
	GGATGGGGTC	CCATTTATTA	TTTACAAATT	CACATATACA	ACAACGCCGT	CCCCĆGTGCC
20	CGCAGTTTTT	ATTAAACATA	GCGTGGGATC	TCCACGCGAA	TCTCGGGTAC.	GTGTTCCGGA
	CATGGGCTCT	TCTCCGGTAG	CGGCGGAGCT	TCCACATCCG	AGCCCTGGTC	CCATGCCTCC
	AGCGGCTCAT	GGTCGCTCGG	CAGCTCCTTG	CTCCTAACAG	TGGAGGCCAG	ACTTAGGCAC
	AGCACAATGC	CCACCACCAC	CAGTGTGCCG	CACAAGGCCG	TGGCGGTAGG	GTATGTGTCT
	GAAAATGAGC	GTGGAGATTG	GGCTCGCACG	GCTGACGCAG	ATGGAAGACT	TAAGGCAGCG
25	GCAGAAGAAG	ATGCAGGCAG	CTGAGTTGTT	GTATTCTGAT	AAGAGTCAGA	GGTAACTCCC
	GTTGCGGTGC	TGTTAACGGT	GGAGGGCAGT	GTAGTCTGAG	CAGTACTCGT	TGCTGCCGCG
	CGCGCCACCA	GACATAATAG	CTGACAGACT	AACAGACTGT	TCCTTTCCAT	GGGTCTTTTC
	TGCAGTCACC	GTCCTTAGAT	CACCATGGAT	GCAATGAAGA	GAGGGCTCTG	CTGTGTGCTG
	CTGCTGTGTG	GAGCAGTCTT	CGTTTCGCCC	AGCGAGATCT	GCTGTGCCTT	CTAGTTGCCA
30	GCCATCTGTT	GTTTGCCCCT	CCCCCGTGCC	TTCCTTGACC	CTGGAAGGTG	CCACTCCCAC
	TGTCCTTTCC	TAATAAAATG	AGGAAATTGC	ATCGCATTGT	CTGAGTAGGT	GTCATTCTAT
	TCTGGGGGGT	GGGGTGGGGC	AGGACAGCAA	GGGGGAGGAT	TGGGAAGACA	ATAGCAGGCA
	TGCTGGGGAT	GCGGTGGGCT	CTATGGCCGC	TGCGGCCAGG	TGCTGAAGAA	TTGACCCGGT
	TCCTCCTGGG	CCAGAAAGAA	GCAGGCACAT	CCCCTTCTCT	GTGACACACC	CTGTCCACGC

	CCCTGGTTCT	TAGTTCCAGC	CCCACTCATA	GGACACTCAT	AGCTCAGGAG	GGCTCCGCCT
	TCAATCCCAC	CCGCTAAAGT	ACTTGGAGCG	GTCTCTCCCT	CCCTCATCAG	CCCACCAAAC
	CAAACCTAGC	CTCCAAGAGT	GGGAAGAAAT	TAAAGCAAGA	TAGGCTATTA	AGTGCAGAGG
	GAGAGAAAAT	GCCTCCAACA	TGTGAGGAAG	TAATGAGAGA	AATCATAGAA	TTTCTTCCGC
5	TTCCTCGCTC	ACTGACTCGC	TGCGCTCGGT	CGTTCGGCTG	CGGCGAGCGG	TATCAGCTCA
	CTCAAAGGCG	GTAATACGGT	TATCCACAGA	ATCAGGGGAT	AACGCAGGAA	AGAACATGTG
	AGCAAAAGGC	CAGCAAAAGG	CCAGGAACCG	TAAAAAGGCC	GCGTTGCTGG	CGTTTTTCCA
	TAGGCTCCGC	CCCCTGACG	AGCATCACAA	AAATCGACGC	TCAAGTCAGA	GGTGGCGAAA
	CCCGACAGGA	CTATAAAGAT	ACCAGGCGTT	TCCCCCTGGA	AGCTCCCTCG	TGCGCTCTCC
10	TGTTCCGACC	CTGCCGCTTA	CCGGATACCT	GTCCGCCTTT	CTCCCTTCGG	GAAGCGTGGC
	GCTTTCTCAT	AGCTCACGCT	GTAGGTATCT	CAGTTCGGTG	TAGGTCGTTC	GCTCCAAGCT
	GGGCTGTGTG	CACGAACCCC	CCGTTCAGCC	CGACCGCTGC	GCCTTATCCG	GTAACTATCG
	TCTTGAGTCC	AACCCGGTAA	GACACGACTT	ATCGCCACTG	GCAGCAGCCA	CTGGTAACAG
	GATTAGCAGA	GCGAGGTATG	TAGGCGGTGC	TACAGAGTTC	TTGAAGTGGT	GGCCTAACTA
15	CGGCTACACT	AGAAGAACAG	TATTTGGTAT	CTGCGCTCTG	CTGAAGCCAG	TTACCTTCGG
	AAAAAGAGTT	GGTAGCTCTT	GATCCGGCAA	ACAAACCACC	GCTGGTAGCG	GTGGTTTTTT
	TGTTTGCAAG	CAGCAGATTA	CGCGCAGAAA	AAAAGGATCT	CAAGAAGATC	CTTTGATCTT
	TTCTACGGGG	TCTGACGCTC	AGTGGAACGA	AAACTCACGT	TAAGGGATTT	TGGTCATGAG
	ATTATCAAAA	AGGATCTTCA	CCTAGATCCT	TTTAAATTAA	AAATGAAGTT	TTAAATCAAT
20	CTAAAGTATA	TATGAGTAAA	CTTGGTCTGA	CAGTTACCAA	TGCTTAATCA	GTGAGGCACC
	TATCTCAGCG	ATCTGTCTAT	TTCGTTCATC	CATAGTTGCC	TGACTCGGGG	GGGGGGGGG
	CTGAGGTCTG	CCTCGTGAAG	AAGGTGTTGC	TGACTCATAC	CAGGCCTGAA	TCGCCCCATC
	ATCCAGCCAG	AAAGTGAGGG	AGCCACGGTT	GATGAGAGCT	TTGTTGTAGG	TGGACCAGTT
	GGTGATTTTG	AACTTTTGCT	TTGCCACGGA	ACGGTCTGCG	TTGTCGGGAA	GATGCGTGAT
25	CTGATCCTTC	AACTCAGCAA	AAGTTCGATT	TATTCAACAA	AGCCGCCGTC	CCGTCAAGTC
	AGCGTAATGC	TCTGCCAGTG	TTACAACCAA	TTAACCAATT	CTGATTAGAA	AAACTCATCG
	AGCATCAAAT	GAAACTGCAA	TTTATTCATA	TCAGGATTAT	CAATACCATA	TTTTTGAAAA
	AGCCGTTTCT	GTAATGAAGG	AGAAAACTCA	CCGAGGCAGT	TCCATAGGAT	GGCAAGATCC
	TGGTATCGGT	CTGCGATTCC	GACTCGTCCA	ACATCAATAC	AACCTATTAA	TTTCCCCTCG
30	TCAAAAATAA	GGTTATCAAG	TGAGAAATCA	CCATGAGTGA	CGACTGAATC	CGGTGAGAAT
	GGCAAAAGCT	TATGCATTTC	TTTCCAGACT	TGTTCAACAG	GCCAGCCATT	ACGCTCGTCA
	TCAAAATCAC	TCGCATCAAC	CAAACCGTTA	TTCATTCGTG	ATTGCGCCTG	AGCGAGACGA
	AATACGCGAT	CGCTGTTAAA	AGGACAATTA	CAAACAGGAA	TCGAATGCAA	CCGGCGCAGG
	AACACTGCCA	GCGCATCAAC	AATATTTTCA	CCTGAATCAG	GATATTCTTC	TAATACCTGG

AATGCTGTTT TCCCGGGGAT CGCAGTGGTG AGTAACCATG CATCATCAGG AGTACGGATA
AAATGCTTGA TGGTCGGAAG AGGCATAAAT TCCGTCAGCC AGTTTAGTCT GACCATCTCA
TCTGTAACAT CATTGGCAAC GCTACCTTTG CCATGTTTCA GAAACAACTC TGGCGCATCG
GGCTTCCCAT ACAATCGATA GATTGTCGCA CCTGATTGCC CGACATTATC GCGAGCCCAT
TTATACCCAT ATAAATCAGC ATCCATGTTG GAATTTAATC GCGGCCTCGA GCAAGACGTT
TCCCGTTGAA TATGGCTCAT AACACCCCTT GTATTACTGT TTATGTAAGC AGACAGTTTT
ATTGTTCATG ATGATATATT TTTATCTTGT GCAATGTAAC ATCAGAGATT TTGAGACACA
ACGTGGCTTT CCCCCCCC CCATTATTGA AGCATTTATC AGGGTTATTG TCTCATGAGC
GGATACATAT TTGAATGTAT TTAGAAAAAT AAACAAATAG GGGTTCCGCG CACATTTCCC
CGAAAAGTGC CACCTGACGT CTAAGAAACC ATTATTATCA TGACATTAAC CTATAAAAAT
AGGCGTATCA CGAGGCCCTT TCGTC (SEQ ID NO:9).

10

15

20

25

30

VIR - Vaccine vector V1R was constructed to obtain a minimum-sized vaccine vector without unneeded DNA sequences, which still retained the overall optimized heterologous gene expression characteristics and high plasmid yields that VIJ and VIJns afford. It was determined that (1) regions within the pUC backbone comprising the E. coli origin of replication could be removed without affecting plasmid yield from bacteria; (2) the 3'-region of the kan<sup>r</sup> gene following the kanamycin open reading frame could be removed if a bacterial terminator was inserted in its place; and, (3) ~300 bp from the 3'- half of the BGH terminator could be removed without affecting its regulatory function (following the original KpnI restriction enzyme site within the BGH element). V1R was constructed by using PCR to synthesize three segments of DNA from V1Jns representing the CMVintA promoter/BGH terminator, origin of replication, and kanamycin resistance elements. respectively. Restriction enzymes unique for each segment were added to each segment end using the PCR oligomers: SspI and XhoI for CMVintA/BGH; EcoRV and BamHI for the kan r gene; and, BelI and SalI for the ori r. These enzyme sites were chosen because they allow directional ligation of each of the PCR-derived DNA segments with subsequent loss of each site: EcoRV and SspI leave blunt-ended DNAs which are compatible for ligation while BamHI and BclI leave complementary overhangs as do SalI and XhoI. After obtaining these segments by PCR each segment was digested with the appropriate restriction enzymes indicated above and then ligated together in a single reaction mixture containing all three DNA segments. The 5'-end of the ori r was designed to include the T2 rho independent terminator sequence that is normally found in this region so that it could provide termination

information for the kanamycin resistance gene. The ligated product was confirmed by restriction enzyme digestion (>8 enzymes) as well as by DNA sequencing of the ligation junctions. DNA plasmid yields and heterologous expression using viral genes within V1R appear similar to V1Jns. The net reduction in vector size achieved was 1346 bp (V1Jns = 4.86 kb; V1R = 3.52 kb). PCR oligomer sequences used to synthesize V1R (restriction enzyme sites are underlined and identified in brackets following sequence) are as follows: (1) 5'-GGTACAAATATTGGCTATTGG CCATTGCATACG-3' (SEQ ID NO:19) [SspI]; (2) 5'-CCACATCTCGAGGAAC CGGGTCAATTCTTCAGCACC-3' (SEQ ID NO:20) [XhoI] (for CMVintA/BGH segment); (3) 5'-GGTACAGATATCGGAAAGCCACGTTGTG TCTCAAAATC-3' 10 (SEQ ID NO:21) [EcoRV]; (4) 5'-CACATGGATCCGTAAT GCTCTGCCAGTGTT ACAACC-3' (SEQ ID NO:2) [BamHI], (for kanamycin resistance gene segment) (5) 5'-GGTACATG ATCACGTAGAAAAGATCA AAGGATCTTCTTG-3' (SEQ ID NO:23) [BcII]; (6) 5'-CCACATGTCGACCCGTAAA AAGGCCGCGTTGCTGG-3' 15 (SEQ ID NO:24): [SalI], (for E. coli origin of replication). The nucleotide sequence of vector V1R is as follows: TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG TCAGGGCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG CTATTGGCCA TTGCATACGT TGTATCCATA TCATAATATG TACATTTATA TTGGCTCATG TCCAACATTA CCGCCATGTT GACATTGATT ATTGACTAGT TATTAATAGT AATCAATTAC GGGGTCATTA GTTCATAGCC CATATATGGA GTTCCGCGTT ACATAACTTA CGGTAAATGG CCCGCCTGGC TGACCGCCCA ACGACCCCCG CCCATTGACG TCAATAATGA CGTATGTTCC 25 CATAGTAACG CCAATAGGGA CTTTCCATTG ACGTCAATGG GTGGAGTATT TACGGTAAAC TGCCCACTTG GCAGTACATC AAGTGTATCA TATGCCAAGT ACGCCCCCTA TTGACGTCAA TGACGGTAAA TGGCCCGCCT GGCATTATGC CCAGTACATG ACCTTATGGG ACTTTCCTAC TTGGCAGTAC ATCTACGTAT TAGTCATCGC TATTACCATG GTGATGCGGT TTTGGCAGTA CATCAATGGG CGTGGATAGC GGTTTGACTC ACGGGGATTT CCAAGTCTCC ACCCCATTGA 30 CGTCAATGGG AGTTTGTTTT GGCACCAAAA TCAACGGGAC TTTCCAAAAT GTCGTAACAA CTCCGCCCA TTGACGCAAA TGGGCGGTAG GCGTGTACGG TGGGAGGTCT ATATAAGCAG AGCTCGTTTA GTGAACCGTC AGATCGCCTG GAGACGCCAT CCACGCTGTT TTGACCTCCA TAGAAGACAC CGGGACCGAT CCAGCCTCCG CGGCCGGGAA CGGTGCATTG GAACGCGGAT TCCCCGTGCC AAGAGTGACG TAAGTACCGC CTATAGAGTC TATAGGCCCA CCCCCTTGGC

	TTCTTATGCA	TGCTATACTG	TTTTTGGCTT	GGGGTCTATA	CACCCCCGCT	TCCTCATGTT
	ATAGGTGATG	GTATAGCTTA	GCCTATAGGT	GTGGGTTATT	GACCATTATT	GACCACTCCC
	CTATTGGTGA	CGATACTTTC	CATTACTAAT	CCATAACATG	GCTCTTTGCC	ACAACTCTCT
	TTATTGGCTA	TATGCCAATA	CACTGTCCTT	CAGAGACTGA	CACGGACTCT	GTATTTTTAC
5	AGGATGGGGT	CTCATTTATT	ATTTACAAAT	TCACATATAC	AACACCACCG	TCCCCAGTGC
	CCGCAGTTTT	TATTAAACAT	AACGTGGGAT	CTCCACGCGA	ATCTCGGGTA	CGTGTTCCGG
	ACATGGGCTC	TTCTCCGGTA	GCGGCGGAGC	TTCTACATCC	GAGCCCTGCT	CCCATGCCTC
	CAGCGACTCA	TGGTCGCTCG	GCAGCTCCTT	GCTCCTAACA	GTGGAGGCCA	GACTTAGGCA
	CAGCACGATG	CCCACCACCA	CCAGTGTGCC	GCACAAGGCC	GTGGCGGTAG	GGTATGTGTC
10	TGAAAATGAG	CTCGGGGAGC	'GGGCTTGCAC	CGCTGACGCA	TTTGGAAGAC	TTAAGGCAGC
	GGCAGAAGAA	GATGCAGGCA	GCTGAGTTGT	TGTGTTCTGA	TAAGAGTCAG	AGGTAACTCC
	CGTTGCGGTG	CTGTTAACGG	TGGAGGGCAG	TGTAGTCTGA	GCAGTACTCG	TTGCTGCCGC
	GCGCGCCACC	AGACATAATA	GCTGACAGAC	TAACAGACTG	TTCCTTTCCA	TGGGTCTTTT
	CTGCAGTCAC	CGTCCTTAGA	TCTGCTGTGC	CTTCTAGTTG	CCAGCCATCT	GTTGTTTGCC
15	CCTCCCCCT	GCCTTCCTTG	ACCCTGGAAG	GTGCCACTCC	CACTGTCCTT	TCCTAATAAA
	ATGAGGAAAT	TGCATCGCAT	TGTCTGAGTA	GGTGTCATTC	TATTCTGGGG	GGTGGGGTGG
	GGCAGCACAG	CAAGGGGGAG	GATTGGGAAG	ACAATAGCAG	GCATGCTGGG	GATGCGGTGG
	GCTCTATGGG	TACCCAGGTG	CTGAAGAATT	GACCCGGTTC	CTCCTGGGCC	AGAAAGAAGC
	AGGCACATCC	CCTTCTCTGT	GACACACCCT	GTCCACGCCC	CTGGTTCTTA	GTTCCAGCCC
20	CACTCATAGG	ACACTCATAG	CTCAGGAGGG	CTCCGCCTTC	AATCCCACCC	GCTAAAGTAC
	TTGGAGCGGT	CTCTCCCTCC	CTCATCAGCC	CACCAAACCA	AACCTAGCCT	CCAAGAGTGG
	GAAGAAATTA	AAGCAAGATA	GGCTATTAAG	TGCAGAGGGA	GAGAAAATGC	CTCCAACATG
	TGAGGAAGTA	ATGAGAGAAA	TCATAGAATT	TCTTCCGCTT	CCTCGCTCAC	TGACTCGCTG
	CGCTCGGTCG	TTCGGCTGCG	GCGAGCGGTA	TCAGCTCACT	CAAAGGCGGT	AATACGGTTA
25	TCCACAGAAT	CAGGGGATAA	ÇGCAGGAAAG	AACATGTGAG	CAAAAGGCCA	GCAAAAGGCC
	AGGAACCGTA	AAAAGGCCGC	GTTGCTGGCG	TTTTTCCATA	GGCTCCGCCC	CCCTGACGAG
	CATCACAAAA	ATCGACGCTC	AAGTCAGAGG	TGGCGAAACC	CGACAGGACT	ATAAAGATAC
	CAGGCGTTTC	CCCCTGGAAG	CTCCCTCGTG	CGCTCTCCTG	TTCCGACCCT	GCCGCTTACC
	GGATACCTGT	CCGCCTTTCT	CCCTTCGGGA	AGCGTGGCGC	TTTCTCAATG	CTCACGCTGT
30	AGGTATCTCA	GTTCGGTGTA	GGTCGTTCGC	TCCAAGCTGG	GCTGTGTGCA	CGAACCCCCC
	GTTCAGCCCG	ACCGCTGCGC	CTTATCCGGT	AACTATCGTC	TTGAGTCCAA	CCCGGTAAGA
	CACGACTTAT	CGCCACTGGC	AGCAGCCACT	GGTAACAGGA	TTAGCAGAGC	GAGGTATGTA
	GGCGGTGCTA	CAGAGTTCTT	GAAGTGGTGG	CCTAACTACG	GCTACACTAG	AAGGACAGTA
	TTTGGTATCT	GCGCTCTGCT	GAAGCCAGTT	ACCTTCGGAA	AAAGAGTTGG	TAGCTCTTGA

TCCGGCAAAC AAACCACCGC TGGTAGCGGT GGTTTTTTTG TTTGCAAGCA GCAGATTACG CGCAGAAAAA AAGGATCTCA AGAAGATCCT TTGATCTTTT CTACGGGGTC TGACGCTCAG TGGAACGAAA ACTCACGTTA AGGGATTTTG GTCATGAGAT TATCAAAAAG GATCTTCACC TAGATCCTTT TAAATTAAAA ATGAAGTTTT AAATCAATCT AAAGTATATA TGAGTAAACT TGGTCTGACA GTTACCAATG CTTAATCAGT GAGGCACCTA TCTCAGCGAT CTGTCTATTT CGTTCATCCA TAGTTGCCTG ACTCCGGGGG GGGGGGGGCGC TGAGGTCTGC CTCGTGAAGA AGGTGTTGCT GACTCATACC AGGCCTGAAT CGCCCCATCA TCCAGCCAGA AAGTGAGGGA GCCACGGTTG ATGAGAGCTT TGTTGTAGGT GGACCAGTTG GTGATTTTGA ACTTTTGCTT TGCCACGGAA CGGTCTGCGT TGTCGGGAAG ATGCGTGATC TGATCCTTCA ACTCAGCAAA 10 AGTTCGATTT ATTCAACAAA GCCGCCGTCC CGTCAAGTCA GCGTAATGCT CTGCCAGTGT TACAACCAAT TAACCAATTC TGATTAGAAA AACTCATCGA GCATCAAATG AAACTGCAAT TTATTCATAT CAGGATTATC AATACCATAT TTTTGAAAAA GCCGTTTCTG TAATGAAGGA GAAAACTCAC CGAGGCAGTT CCATAGGATG GCAAGATCCT GGTATCGGTC TGCGATTCCG ACTCGTCCAA CATCAATACA ACCTATTAAT TTCCCCTCGT CAAAAATAAG GTTATCAAGT 15 GAGAAATCAC CATGAGTGAC GACTGAATCC GGTGAGAATG GCAAAAGCTT ATGCATTTCT TTCCAGACTT GTTCAACAGG CCAGCCATTA CGCTCGTCAT CAAAATCACT CGCATCAACC AAACCGTTAT TCATTCGTGA TTGCGCCTGA GCGAGACGAA ATACGCGATC GCTGTTAAAA GGACAATTAC AAACAGGAAT CGAATGCAAC CGGCGCAGGA ACACTGCCAG CGCATCAACA ATATTTTCAC CTGAATCAGG ATATTCTTCT AATACCTGGA ATGCTGTTTT CCCGGGGATC 20 GCAGTGGTGA GTAACCATGC ATCATCAGGA GTACGGATAA AATGCTTGAT GGTCGGAAGA GGCATAAATT CCGTCAGCCA GTTTAGTCTG ACCATCTCAT CTGTAACATC ATTGGCAACG CTACCTTTGC CATGTTTCAG AAACAACTCT GGCGCATCGG GCTTCCCATA CAATCGATAG ATTGTCGCAC CTGATTGCCC GACATTATCG CGAGCCCATT TATACCCATA TAAATCAGCA TCCATGTTGG AATTTAATCG CGGCCTCGAG CAAGACGTTT CCCGTTGAAT ATGGCTCATA ACACCCCTTG TATTACTGTT TATGTAAGCA GACAGTTTTA TTGTTCATGA TGATATATTT TTATCTTGTG CAATGTAACA TCAGAGATTT TGAGACACAA CGTGGCTTTC CCCCCCCCC CATTATTGAA GCATTTATCA GGGTTATTGT CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC GAGGCCCTTT 30 CGTC (SEQ ID NO:25).

#### **EXAMPLE 2**

Codon Optimized HIV-1 Pol and HIV-1 IA Pol Derivatives as DNA Vector Vaccines Synthesis of WT-optpol and IA-opt-pol Gene - Construction of both genes were conducted by Midland Certified Reagent Company (Midland, TX) following established strategies. Ten double stranded oligonucleotides, ranging from 159 to 340 bases long and encompassing the entire pol gene, were synthesized by solid state methods and cloned separately into pUC18. For the wt-pol gene, the fragments are as follows:

	BgIII#1-Ecl136II half site at 282	= pJS6A1-7
10	PmII half site at #285 - Ecl136II half site at #597	= pJS6B2-5
	SspI half site at #600 - Ecl136II half site at #866	= pJS6C1-4
	SmaI half site at #869 – ApaI #1095	= pJS6D1-4
	ApaI #1095 - KpnI #1296	= pJS6E1-4
	KpnI #1296 – XcmI #1636	= pJS6F1-5
15	XcmI #1636 – NsiI #1847	= pJS6G1-2
	<i>Nsi</i> I #1847 – <i>BcI</i> I half site at #2174	= pJS6H1-14
	BcII half site at #2174 – SacI #2333	= pJS6I1-2
	SacI #2333 - BglII #2577	= pJS6J1-1

5

20

25

EcoRI and HindIII sequences were added upstream of each 5' end and downstream of each 3' end, respectively, to allow cloning into the EcoRI-HindIII sites of pUC18.

The next stage of the synthesis was to consolidate these cassettes into three roughly equal fragments (alpha, beta, gamma) and was performed as follows:

Alpha: The SspI-HindIII small fragment of pJS6C1-4 was transferred into the Ecl136II-HindIII sites of pJS6B2-5 to give pJS6BC1-1. Into the EcoRI-PmII sites of this plasmid was inserted the EcoRI-Ecl136II small fragment of pJS6A1-7 to give pJS6 $\alpha$ 1-8.

Beta: The *Eco*RI-*Apa*I small fragment of pJS6D1-4 was inserted into the corresponding sites of pJS6E1-2 to give pJS6DE1-2. Also, the *Eco*RI-*Xcm*I small fragment of pJS6F1-5 was inserted into the corresponding sites of pJS6G1-2 to give pJS6FG1-1. Then the *Eco*RI-*Kpn*I small fragment of pJS6DE1-2 was inserted into the corresponding sites of pJS6FG1-1 to give pJS6β1-1.

Gamma: The SacI-HindIII small fragment of pJS6J1-1 was inserted into the corresponding sites of pJS6I1-2 to give pJS6IJ1-1. This plasmid was propagated through E. coli SCS110 (dam-/dcm-) to permit subsequent cleavage at the BcII site.

The BcII-HindIII small fragment of the unmethylated pJS6IJ1-1 was inserted into the BgIII-HindIII sites of pJS6H1-14 to give pJS6x1-1.

The wt-pol alpha, beta, gamma were ligated into the entire sequence as follows:

5 The *Eco*RI-*Ecl*136II small fragment of pJS6α1-8 was inserted into the *Eco*RI-*Sma*I sites of pJS6β1-1 to give pJS6αβ2-1.

Into the NsiI-HindIII sites of this plasmid was inserted the NsiI-HindIII small fragment of pJS6 $\chi$ 1-1 to give pUC18-wt-pol. This final plasmid was completely resequenced in both strands.

To construct the entire IA-pol gene, only 3 new small fragments were synthesized:

PmII half site at #285 – Ecl136II half site at #597 = pJS7B1-1

KpnI #1296 - XcmI #1636 = pJS7F1-2

NsiI #1847 - BgIII half site at #2174 = pJS7H1-5

These were then used in the same reconstruction strategy as described above to give pUC18-IA-pol.

20

Expression Vector Construction - pUC18-wt-pol and pUC18-IA-pol were digested with BgIII in order to isolate fragments containing the entire pol genes. V1R, V1Jns, V1Jns-tpa (Shiver, et al., 1995, Immune responses to HIV gp120 elicited by DNA vaccination. In Vaccines 95 (eds. Chanock, R. M., Brown, F., Ginsberg, H.S., & Norrby, E.) @ pp. 95-98; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; see also Example Section 1) were digested with BgIII. The cut vectors were then treated with calf intestinal alkaline phosphatase. Both wt-pol and IA-pol genes were ligated into cut V1R using T4 DNA ligase (16 °C, overnight).

Competent DH5α cells were transformed with aliquots of the ligation mixtures.
 Colonies were screened by restriction digestion of amplified plasmid isolates.
 Following a similar strategy, the BgIII fragment containing the IA-pol was subcloned into the BgIII site of V1Ins. To ligate the IA-pol gene into V1Ins-tpa, the IA-pol gene was PCR-amplified from V1R-IA-pol using pfu polymerase and the following pair of primers: 5'-GGTACAAGATCTCCGCCCCATCTCCCCCATTGAGA-3'
 (SEQ ID NO:26), and 5'-CCACATAGATCTGCCCGGGCTTTAGTCCTCATC-3'

(SEQ ID NO:27). The upstream primer was designed to remove the initiation met codon and place the pol gene in frame with the tpa leader coding sequence from

V1Jns-tpa. The PCR product was purified from the agarose gel slab using Sigma

DNA Purification spin columns. The purified products were digested with BgIII and subcloned into the BgIII site of V1Jns-tpa.

Results - The codon humanized wt- and IA-pol genes were constructed via stepwise ligation of 10 synthetic dsDNA fragments (Ferretti, et al., 1986, Proc. Natl. Acad. Sci. USA 83: 599-603). For expression in mammalian systems, the IA-pol gene was subcloned into V1R, V1Jns, and V1Jns-tpa. All these vectors place the gene under the control of the human cytomegalovirus/intron A hybrid promoter (hCMVIA). The DNA sequence of the IA-pol gene and the expressed protein product are shown in Figure 2A-B. Subcloning into V1Jns-tpa attaches the leader sequence from human tissue-specific plasminogen activator (tpa) to the N-terminus of the IA-pol (Pennica, et al., 1983, Nature 301: 214-221) to allow secretion of the protein. The sequences of the tpa leader and the fusion junction are shown in Figure 3.

### **EXAMPLE 3**

15

10

#### HIV-1 POL Vaccine - Rodent Studies

Materials - E. coli DH5α strain, penicillin, streptomycin, ACK lysis buffer, hepes, L-glutamine, RPMI1640, and ultrapure CsCl were obtained from Gibco/BRL (Grand Island, NY). Fetal bovine serum (FBS) was purchased from Hyclone. Kanamycin, Tween 20, bovine serum albumin, hydrogen peroxide (30%), 20 concentrated sulfuric acid,  $\beta$ -mercaptoethanol ( $\beta$ -ME), and concanavalin A were obtained from Sigma (St. Louis, MO). Female balb/c mice at 4-6 wks of age were obtained from Taconic Farms (Germantown, NY). 0.3-mL insulin syringes were purchased from Myoderm. 96-well flat bottomed Maxisorp plates were obtained form NUNC (Rochester, NY). HIV-1<sub>IIIB</sub> RT p66 recombinant protein was obtained from 25 Advanced Biotechnologies, Inc. (Columbia, MD). 20-mer peptides were synthesized by Research Genetics (Huntsville, AL). Horseradish peroxidase (HRP)-conjugated rabbit anti-mouse IgG1 was obtained from ZYMED (San Francisco, CA). 1,2phenylenediamine dihydrochloride (OPD) tablets was obtained from DAKO (Norway). Purified rat anti-mouse IFN-gamma (IgG1, clone R4-6A2), biotin-30 conjugated rat anti-mouse IFN-gamma (IgG1, clone XMG 1.2), and strepavidinalkaline phosphatase conjugate were purchased from PharMingen (San Diego, CA). 1-STEP NBT/BCIP dye was obtained from Pierce Chemicals (Rockford, IL). 96-well Multiscreen membrane plate was purchased from Millipore (France). Cell strainer was obtained from Becton-Dickinson (Franklin Lakes, NJ).

Plasmid Preparation - E. coli DH5α cells expressing the pol plasmids were grown to saturation in LB broth supplemented with 100 ug/mL kanamycin. Plasmid were purified by standard CsCl method and solubilized in saline at concentrations greater than 5 mg/mL until further use.

Vaccination - The plasmids were prepared in phosphate-buffered saline and administered into balb/c by needle injection (28-1/2G insulin syringe) of 50 uL aliquot into each quad muscle. V1Jns-IApol was administered at 0.3, 3, 30 ug dose and for comparison, V1Jns-tpa-IApol was given at 30 ug dose. Immunizations were conducted at T=0 and T=8 wks (for select animals from the 30-ug dose cohorts).

5

10

15

20

25

30

ELISA Assay - At T=12 wks, blood samples were collected by making an incision of a tail vein and the serum separated. Anti-RT titers were obtained following standard secondary antibody-based ELISA. Briefly, Maxisorp plates were coated by overnight incubation with 100 uL of 1 ug/mL HIV-1 RT protein (in PBS). The plates were washed with PBS/0.05% Tween 20 and incubated for approx. 2h with 200 uL/well of blocking solution (PBS/0.05% tween/1% BSA). The blocking solution was decanted; 100 uL aliquot of serially diluted serum samples were added per well and incubated for 2 h at room temperature. The plates were washed and 100 uL of 1/1000-diluted HRP-rabbit anti-mouse IgG were added with 1 h incubation. The plates were washed thoroughly and soaked with 100 uL OPD/H<sub>2</sub>O<sub>2</sub> solution for 15 min. The reaction was quenched by adding 100 uL of 0.5M H<sub>2</sub>SO4 per well. OD<sub>492</sub> readings were recorded.

ELIspot - Spleens were collected from 5 mice/cohort at T=13-14 wks and pooled into a tube of 8-mL R10 medium (RPMI1640, 10% FBS, 2mM L-glutamine, 100U/mL Penicillin, 100 u/mL streptomycin, 10 mM Hepes, 50 uM β-ME). Multiscreen opaque plates were coated with 100μl/well of capture mAb (purified R4-6A2 diluted in PBS to 5μg/ml) at 4°C overnight. The plates were washed with PBS/Pen/Strep in hood and blocked with 200μl/well of complete R10 medium for 37°C for at least 2 hrs. The mouse spleens were ground on steel mesh, collected into 15ml tubes and centrifuged at 1200rpm for 10min. The pellet was treated in ACK buffer (4ml of lysis buffer per spleen) for 5min at room temperature to lyse red blood cells. The cell pellet was centrifuged as before, resuspended in K-medium (5ml per mouse spleen), filtered through a cell strainer and counted using a hemacytometer. Block medium was decanted from the plates and 100μl/well of cell samples (5.0x10e5 cells per well) plus antigens were added. Pol-specific CD4<sup>+</sup> cells were stimulated

using a mixture of previously identified two epitope-containing peptides (aa641-660, aa731-750). Antigen-specific CD8+ cells were stimulated using a pool of four peptide epitope-containing peptides (aa201-220, aa311-330, aa571-590, aa781-800) or with individual peptides. A final concentration of 4 ug/mL per peptide was used. Each splenocyte sample is tested for IFN-gamma secretion by adding the mitogen, concanavalin A. Plates were incubated at 37°C, 5% CO<sub>2</sub> for 20-24 h. The plates were washed with PBS/0.05% Tween 20 and soaked with 100 uL/well of 5 ug/mL biotin-conjugated rat anti-mouse IFN- mAb (clone XMG1.2) at 4°C overnight. The plates were washed and soaked with 100 uL/well 1/2500 dilution of strepavidin-AP (in PBS/0.005% Tween/5%FCS) for 30 min at 37 °C. Following a wash, spots were developed by incubating with 100µl/well 1-step NBT/BCIP for 6-10 min. The plates were washed with water and allowed to air dry. The number of spots in each wells were determined using a dissecting microscope and normalized to 10e6 cells.

10

15

20

25

Results - Single vaccination of balb/c mice with V1Jns-IApol is able to induce antigen-specific antibody (Figure 4) and T cell (Figure 5) responses in a dose response manner. IFN-gamma secretion from splenocytes can be detected from 3 and 30 ug cohort following stimulation with pools of peptides that contain CD4+ and CD8+ T cell epitopes. These epitopes were identified by (1) screening 20-mer peptides that encompass the entire pol sequence and overlap by 10 amino acid for ability to stimulate IFN-gamma secretion from vaccinee splenocytes, and (2) determining the T cell type (CD4+ or CD8+) by depleting either population in an Elispot assay. Addition of tpa leader sequence to the pol gene is able to induce comparable, if not slightly higher, frequencies of pol-specific CD4+ and CD8+ cells. A second immunization with either V1Jns-IApol and V1Jns-tpa-IApol resulted in effective boosting of the immune responses.

#### **EXAMPLE 4**

HIV-1 Pol Vaccine - Non Human Primate Studies

Materials - E. coli DH5α strain, penicillin, streptomycin, and ultrapure CsCl
were obtained from Gibco/BRL (Grand Island, NY). Kanamycin and
phytohemagluttinin (PHA-M) were obtained from Sigma (St. Louis, MO). 20-mer
peptides were synthesized by SynPep (Dublin, CA) and Research Genetics
(Huntsville, AL). 96-well Multiscreen Immobilon-P membrane plates were obtained
from Millipore (France). Strepavidin-alkaline phosphatase conjugate were purchased

form Pharmingen (San Diego, CA). 1-Step NBT/BCIP dye was obtained form Pierce Chemicals (Rockford, IL). Rat anti-human IFN-gamma mAb and biotin-conjugated anti-human IFN-gamma reagent were obtained from R&D Systems (Minneapolis, MN). Dynabeads M-450 anti-human CD4 were obtained from Dynal (Norway). HIVp24 antigen assay was purchased from Coulter Corporation (Miami, FL). HIV-1<sub>IIIB</sub> RT p66 recombinant protein was obtained from Advanced Biotechnologies, Inc. (Columbia, MD). Plastic 8 well strips/plates, flat bottom, Maxisorp, are obtained from NUNC (Rochester, NY). HIV+ human serum 9711234 was obtained from Biological Specialty Corp.

Plasmid Preparation - E. coli DH5α cells expressing the pol plasmids were grown to saturation in LB supplemented with 100 ug/mL kanamycin. Plasmid were purified by standard CsCl method and solubilized in saline at concentrations greater than 5 mg/mL until further use.

10

15

20

25

30

Vaccination - Cohorts of 3 rhesus macaques (approx. 5-10 kg) were vaccinated with 5 mg dose of either V1Jns-IApol or V1Jns-tpa-IApol. The vaccine was administered by needle injection of two 0.5 mL aliquots of 5 mg/mL plasmid solution (in phosphate-buffered saline, pH 7.2) into both deltoid muscles. Prior to vaccination, the monkeys were chemically restraint with i.m. injection of 10 mg/kg ketamine. The animals were immunized 3x at 4 week intervals (T=0, 4, 8 wks).

Sample Collection - Blood samples were collected at T = 0, 4, 8, 12, 16, 18 wks; sera and PBMCs were isolated using established protocols.

ELIspot Assay - Immobilon-IP plates were coated with 100 uL/well of rat antihuman IFN-gamma mAb at 15 ug/mL at 4 °C overnight. The plates are then washed with PBS and block by adding 200 uL/well of R10 medium. 4x10e5 peripheral blood cells were plated per well and to each well, either media or one of the pol peptide pools (final concentration of 4 ug/mL per peptide) or PHA, a known mitogen, is added to a final volume of 100 uL. Duplicate wells were set up per sample per antigen and stimulation was performed for 20-24 h at 37 °C. The plates are then washed; biotinylated anti-human IFN-gamma reagent is added (0.1 ug/mL, 100 uL per well) and allowed to incubate for overnight at 4 °C. The plates are again washed and 100 uL of 1:2500 dilution of the strepavidin-alkaline phosphatase reagent (in PBS/0.005% Tween/5% FCS) is added and allowed to incubate for 2 h at ambient room temperature. After another wash, spots are developed by incubating with 100 uL/well of 1-step NBT/BCIP for 6-10 min. CD4-T cell depletion was performed by

adding 1 bead particle/10 cell of Dynabeads M450 anti-human CD4, prewashed with PBS, and incubating on the shaker at 4  $^{\circ}$ C for 30 min. The beads are fractionated magnetically and the unbound cells collected and quantified before plating onto the ELISpot assay plates (at  $4\times10e5$  cells per well).

5

10

15

20

25

30

CTL Assay - Procedures for establishing bulk CTL culture with fresh or cryopreserved peripheral blood mononuclear cells (PBMC) are as follows. Twenty percent total PBMC were infected in 0.5 ml volume with recombinant vaccinia virus, Vac-tpaPol, respectively, at multiplicity of infection (moi) of 5 for 1 hr at 37°C, and then combined with the remaining PBMC sample. The cells were washed once in 10 ml R-10 medium, and plated in a 12 well plate at approximately 5 to 10 x 10<sup>6</sup> cells/well in 4 ml R-10 medium. Recombinant human IL-7 was added to the culture at the concentration of 330 U/ml. Two or three days later, one milliliter of R-10 containing recombinant human IL-2 (100 U/ml) was added to each well. And twice weekly thereafter, two milliliters of cultured media were replaced with 2 ml fresh R-10 medium with rhIL-2 (100 U/ml). The lymphocytes were cultured at 37°C in the presence of 5% CO<sub>2</sub> for approximately 2 weeks, and used in cytotoxicity assay as described below. The effector cells harvested from bulk CTL cultures were tested against autologous B lymphoid cell lines (BLCL) sensitized with peptide pools. To prepare for the peptide-sensitized targets, the BLCL cells were washed once with R-10 medium, enumerated, and pulsed with peptide pool (about 4 to 8 ug/ml concentration for each individual peptide) in 1 ml volume overnight. A mock target was prepared by pulsing cells with peptide-free DMSO diluent to match the DMSO concentration in the peptide-pulsed targets. The cells were enumerated the next morning, and 1 x 10<sup>6</sup> cells were resuspended in 0.5 ml R-10 medium. Five to ten microliters of Na 51 CrO<sub>4</sub> were added to the tubes at the same time, and the cells were incubated for 1 to 2 hr 37°C. The cells were then washed 3 times and resuspended at 5x10<sup>4</sup> cells/ml in R-10 medium to be used as target cells. The cultured lymphocytes were plated with target cells at designated effector to target (E:T) ratios in triplicates in 96-well plates, and incubated at 37°C for 4 hours in the presence of 5% CO<sub>2</sub>. A sample of 30  $\mu$ l supernatant from each well of cell mixture was harvested onto a well of a Lumaplate-96 (Packard Instrument, Meriden, CT), and the plate was allowed to air dry overnight. The amount of <sup>51</sup>Cr in the well was determined through betaparticle emission, using a plate counter from Packard Instrument. The percentage of specific lysis was calculated using the formula as: % specific lysis = (E-S) / (M-S).

The symbol E represents the average cpm released from target cells in the presence of effector cells, S is the spontaneous cpm released in the presence of medium only, and M is the maximum cpm released in the presence of 2% Triton X-100.

5

10

15

20

25

ELISA Assay - The pol-specific antibodies in the monkeys were measured in a competitive RT EIA assay, wherein sample activity is determined by the ability to block RT antigen from binding to coating antibody on the plate well. Briefly, Maxisorp plates were coated with saturating amounts of pol positive human serum (97111234). 250 uL of each sample is incubated with 15 uL of 266 ng/mL RT recombinant protein (in RCM 563, 1% BSA, 0.1% tween, 0.1% NaN<sub>3</sub>) and 20 uL of lysis buffer (Coulter p24 antigen assay kit) for 15 min at room temperature. Similar mixtures are prepared using serially diluted samples of a standard and a negative control which defines maximum RT binding. 200 uL/well of each sample and standard were added to the washed plate and the plate incubated 16-24 h at room temperature. Bound RT is quantified following the procedures described in Coulter p24 assay kit and reported in milliMerck units per mL arbitrarily defined by the chosen standard.

Results - Repeated vaccinations with V1Jns-IApol induced in 1 of 3 monkeys (94R033) significant levels of antigen-specific T cell activation (Figure 6A-C and Table 2) and CTL killing of peptide-pulsed autologous cells (Figure 7A-B). A significant CD8+ component to the T cell responses in this animal was confirmed by peptide-stimulation of CD4-depleted PBMCs in an ELIspot assay (Table 2).

Immunization with V1Jns-tpa-IApol produced T cell responses from all 3 vaccinees (Figures 6A-C, Figure 7A-B; Table 2). Two (920078, 94R028) exhibited bulk CTL activity and detectable CD8+ components as measured by Elispot analyses of CD4-depleted PBMCs. For the third monkey (920073), the activated T cells were largely CD4+ (Table 2). Table 3 shows the time course data on the frequency of IFN-gamma secreting cells (SFC/million cells) upon antigen-specific stimulation for monkeys vaccinated 3x with either V1Jns-IApol or V1Jns-tpa-IApol (5 mg dose). At T=18 wks, CD4-cell depletion were performed; the reported values are the number of spots per million of fractionated cells and are not corrected for the resultant enrichment of CD8+ T cells. PBMCs were stimulated with peptide pools that represent either IA pol protein (mpol-1, mpol-2) or wt Pol (wtpol-1, wtpol-2).

TABLE 2

Vaccine	Animal No.	Antigen	T=0wk	T=4Wk	T=8Wk	T=12Wk	T=1	8Wk
			Dose1	Dose 2	Dose 3			CD4Dept
VIJns-IApd 5 mgs	9417008	medum mpd-1 mpd-2 wlpd-1 wlpd-2	1 3 0	15 69 25 49 34	6 28 21 20 24	11 61 19 53 24	11 20 28 18 19	11 15 16
	9417013	mectum mpd-1 mpd-2 wtpd-1 wtpd-2	0 0 1	14 9 15 9 6	6 63 24 50 21	۶ 25 36 33 28	18 34 24 18 25	11 9 15
	9412033	rrectum rrpod-1 rrpod-2 wtpod-1 wtpod-2	4 3 0	15 29 24 30 48	11 86 25 38 46	14 51 43 60 86	13 41 59 53 61	8 24 64
VIJrs-tpo-IApd 5 mgs	920078	medum mpd-1 mpd-2 wtpd-1 wtpd-2	0 3 · 1	24 110 221 115 218	13 120 130 53 204	11 119 561 70 490	14 155 289 116 194	11 11 145
	920073	medum mpd-1 mpd-2 wlpd-1 wlpd-2	0 0 0	13 36 29 20 25	3 51 16 35 16	15 113 83 100 79	15 90 115 74 61	6 14 : 34
•	94R028	rredum mpd-1 mpd-2 wtpd-1 wtpd-2	0 1 1	18 30 24 23 26	11 24 23 25 28	18 29 66 34 71	19 30 59 29 40	9 28 95
Näve	920072	medum mpd-1 mpd-2 wtpd-1 wtpd-2	1 0 1	19 24 24 18 23	3 11 5 13 14	38 25 28 20 33	9 4 6 6 14	4 6 5

For the Elispot assay, antigen specific stimulation were performed by using pools of 20-mer peptide pools based on the vaccine sequence. The vaccine pol sequence differs from the wild-type HIV-1 sequence by 9 point mutations, thereby affecting 16 of the 20-mer peptides in the pool. Comparable responses were observed in the vaccinees when these peptides are replaced with those using the wild-type sequences.

Four of the vaccinees gave anti-RT titers above background after 3 dosages of the plasmids (Table 2).

10

5

TABLE 3
Anti-RT levels in Rhesus Macaques Vaccinated 3x (4 week intervals) with 5 mgs of V1Jns-IApol or V1Jns-tpa-IApol expressed in mMU/mL.

Vaccine/Monkey	T=0Wk	T=4	T=8	T=12	T=16
	DOSE 1	DOSE 2	DOSE 3		
VIJns-IApol, 5 mg					
94R008	ND	<10	<10	15	14
94R013	ND	<10	<10	<10	<10
94R033	ND	< <u>10</u>	<10	25	19
VIJns-tpa-lApol, 5 mg					
920078	3	<b>0</b> [>	<10	35	17.
920073	3	<10	<10	<10	· <10
9417028	ND	<10	<10	20	63

15

#### EXAMPLE 5

# Effect of Codon Optimization on In Vivo Expression and Cellular Immune Response of wt-pol

Materials and Methods - Extraction of virus-derived pol gene - The gene for RT-IN

(wt-pol; a non-codon optimized wild type pol gene derived directly from the HIV IIIB genome) was extracted and amplified from the HIV IIIB genome using two primers,

5'-CAG GCG AGA TCT ACC ATG GCC CCC ATT AGC CCT ATT GAG ACT GTA-3' (SEQ ID NO:29) and 5'-CAG GCG AGA TCT GCC CGG GCT TTA ATC CTC ATC CTG TCT ACT TGC CAC-3' (SEQ ID NO:30), containing BglII sites.

The reaction contained 200 nmol of each primer, 2.5 U of pfu Turbo DNA polymerase (Stratagene, La Jolla, CA), 0.2 mM of each dNTPs, and the template DNA in 10mM KCl, 10mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 20mM Tris-HCl pH 8.75, 2mM MgSO<sub>4</sub>, 0.1% TritonX-100, 0.1mg/ml bovine serum albumin (BSA). Thermocycling

conditions were as follows: 20 cycles of 1 min at 95 °C, 1 min at 56 °C, and 4 mins at 72 °C with 15-min capping at 72 °C. The digested PCR fragment was subcloned into the BgIII site of the expression plasmid V1Ins (Shiver, et al., 1995, Immune responses to HIV gp120 elicited by DNA vaccination. In Chanock, R. M., Brown, F., Ginsberg, H. S., and Norrby, E. (Eds.) Vaccines 95. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp 95-98; see also Example section 1 herein) expression plasmid following similar procedures as described above. The ligation mixtures were then used to transform competent E. coli DH5 cells and screened by PCR amplification of individual colonies. Sequence of the entire gene insert was confirmed. All plasmid constructs for animal immunization were purified by CsCl method (Sambrook, et al., 1989, Fritsch and Maniatis, T. (Eds) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor).

10

15

20

25

In vitro expression in mammalian cells - 1.5x10<sup>6</sup> 293 cells were transfected with 1 or 10 μg of V1R-wt-pol (codon optimized) and V1Jns-wt-pol (virus derived) using the Cell Phect kit and incubated for 48 h at 37 °C, 5% CO<sub>2</sub>, 90% humidity. Supernatants and cell lysates were prepared and assayed for protein content using Pierce Protein Assay reagent (Rockford, IL). Aliquots containing equal amounts of total protein were loaded unto 10-20% Tris glycine gel (Novex, San Diego, CA) along with the appropriate molecular weight markers. The pol product was detected using anti-serum from a seropositive patient (Scripps Clinic, San Diego, CA) diluted 1:1000 and the bands developed using goat anti-human IgG-HRP (Bethyl, Montgomery, TX) at 1:2000 dilution and standard ECL reagent kit (Pharmacia LKB Biotechnology, Uppsala, Sweden).

Ultrasensitive RT activity assay of pol constructs - RT activities from codon optimized wt-pol and IA pol plasmids were analyzed by the Product-Enhanced Reverse Transcriptase (PERT) assay using Perkin Elmer 7700, Taqman technology (Arnold, et al., 1999, One-step fluorescent probe product-enhanced reverse transcriptase assay. In McClelland, M., Pardee, A. (Eds.) Expression genetics: accelerated and high-throughput methods. Biotechniques Books, Natick, MA, pp. 201-210). Background levels for this assay were determined using 1:100,000 dilution of lysates from mock (chemical treatment only, no vector) transfected 293 cells. This background range is set as RT/reaction tube of 0.00 to 56.28 which is taken from the mean value of 13.80 +/- 3 standard deviations (sd=14.16). Any individual value >56.28 would be considered positive for PERT assay. Cells lysates were prepared

similarly for the following samples: mock transfection with empty V1Jns vector; no vector control; transfection with V1Jns-tpa-pol (codon optimized); and transfection with V1Jns-IApol (codon optimized). Samples were serially diluted to 1:100,000 in PERT buffer and 24 replicates for each sample at this dilution were assayed for RT activity.

5

10

15

20

25

30

Rodent immunization with optimized and virus-derived pol plasmids - To compare the immunogenic properties of wt-pol (codon optimized) and virus-derived pol gene, cohorts of BALB/c mice (N=10) were vaccinated with 1 µg, 10 µg, and 100 µg doses of V1R-wt-pol (codon optimized) and V1Jns-wt-pol plasmid (virus derived). At 5 weeks post dose 1, 5 of 10 mice per cohort were boosted with the same dose of plasmid they initially received. In all cases, the vaccines were suspended or diluted in 6 mM sodium phosphate, 150 mM sodium chloride, pH 7.2, and the total dose was injected to both quadricep muscles in 50 µL aliquots using a 0.3-mL insulin syringe with 28-1/2G needles (Becton-Dickinson, Franklin Lakes, NJ).

Anti-RT ELISA - Anti-RT titers were obtained following standard secondary antibody-based ELISA. Maxisorp plates (NUNC, Rochester, NY) were coated by overnight incubation with 100 µL of 1 µg /mL HIV-1 RT protein (Advanced Biotechnologies, Columbia, MD) in PBS. The plates were washed with PBS/0.05% Tween 20 using Titertek MAP instrument (Hunstville, AL) and incubated for approximately 2h with 200 µL/well of blocking solution (PBS/0.05% tween/1%) BSA). The blocking solution was decanted; 100 µL aliquot of serially diluted serum samples were added per well and incubated for 2 h at room temperature. An initial dilution of 100-fold is performed followed by 4-fold serial dilution. The plates were washed and 100 µL of 1/1000-diluted HRP-rabbit anti-mouse IgG (ZYMED, San Francisco, CA) were added with 1 h incubation. The plates were washed thoroughly and soaked with 100 µL 1,2-phenylenediamine dihydrochloride/hydrogen peroxide (DAKO, Norway) solution for 15 min. The reaction was quenched by adding 100 μL of 0.5M H<sub>2</sub>SO4 per well. OD<sub>492</sub> readings were recorded using Titertek Multiskan MCC/340 with S20 stacker. Endpoint titers were defined as the highest serum dilution that resulted in an absorbance value of greater than or equal to 0.1 OD<sub>492</sub> (2.5 times the background value).

ELIspot assay - Antigen-specific INFγ-secreting cells from mouse spleens were detected using the ELIspot assay (Miyahira, et al., 1995, Quantification of antigen specific CD8<sup>+</sup> T cells using an ELISPOT assay. *J. Immunol. Methods* 1995,

5

10

15

20

25

30

181, 45-54). Typically, spleens were collected from 3-5 mice/cohort and pooled into a tube of 8-mL complete RPMI media (RPMI1640, 10% FBS, 2mM L-glutamine, 100U/mL Penicillin, 100 u/mL streptomycin, 10 mM Hepes, 50 uM β-ME). Multiscreen opaque plates (Millipore, France) were coated with 100 µL/well of 5 μg/mL purified rat anti-mouse IFN-γ IgG1, clone R4-6A2 (Pharmingen, San Diego, CA), in PBS at 4°C overnight. The plates were washed with PBS/penicillin/streptomycin in hood and blocked with 200 µL/well of complete RPMI media for 37 °C for at least 2 h. The mouse spleens were ground on steel mesh, collected into 15ml tubes and centrifuged at 1200rpm for 10 min. The pellet was treated with 4 mL ACK buffer (Gibco/BRL) for 5 min at room temperature to lyse red blood cells. The cell pellet was centrifuged as before, resuspended in complete RPMI media (5 ml per mouse spleen), filtered through a cell strainer and counted using a hemacytometer. Block media was decanted from the plates and to each well, 100 μL of cell samples (5x10<sup>5</sup> cells per well) and 100 μL of the antigen solution were added. To the control well, 100 µL of the media were added; for specific responses, peptide pools containing either CD4<sup>+</sup> or CD8<sup>+</sup> epitopes were added. In all cases, a final concentration of 4 µg/mL per peptide was used. Each sample/antigen mixture were performed in triplicate wells. Plates were incubated at 37°C, 5% CO<sub>2</sub>, 90% humidity for 20-24 h. The plates were washed with PBS/0.05% Tween 20 and incubated with 100 µL/well of 1.25 µg/mL biotin-conjugated rat antimouse IFN-γ mAb, clone XMG1.2 (Pharmingen) at 4°C overnight. The plates were washed and incubated with 100 μL/well 1/2500 dilution of strepavidin-alkaline phosphatase conjugate (Pharmingen) in PBS/0.005% Tween/5% FBS for 30 min at 37 °C. Following a wash, spots were developed by incubating with 100 μl/well 1-step NBT/BCIP (Pierce Chemicals) for 6-10 min. The plates were washed with water and allowed to air dry. The number of spots in each well was determined using a dissecting microscope and the data normalized to 10<sup>6</sup> cell input.

Results - In vitro expression of Pol in mammalian cells - Heterologous expression of the optimized wt or IA pol genes (V1R-wt-pol (codon optimized), V1Jns-IApol (codon optimized), V1Jns-IApol (codon optimized)) in 293 cells (Figure 8) yielded a single polypeptide of correct approximate molecular size (90-kDa) for the RT-IN fusion product. In contrast, no expression could be detected by transfecting cells with 1 and 10 μg of the V1Jns-wt-pol, which bears the virus-derived pol.

Ultrasensitive RT assay of cells transfected with Pol constructs - Table 4 summarizes the levels of polymerase activity from mock (vector only) control, IApol (codon optimized)and wt-pol plasmids (codon optimized). Results indicate that the wild-type POL transfected cells contained RT activity approximately 4-5 logs higher than the 293 cell only baseline values. Mock transfected cells contained activity no higher than baseline values. The RT activity from opt-IApol-transfected cells was also found to be no different than baseline values; no individual reaction tube resulted in RT activity higher than the established cut-off value of 56.

10

15

20

5

Table 4

Sample	Avg. RT/tube	Standard deviation	Minimum	Maximum
Vector only	16.25	18.52	0.0	42.99
IApol (codon	2.99	8.01	0.0	35.20
optimized)				·
Wt-pol	126147	21338	68973	152007
(codon				. ,
optimized)				

Comparative immunogenicity of optimized and virus-derived pol plasmid - To compare the *in vivo* potencies of both constructs, BALB/c mice (N=10 per group) were vaccinated with escalating doses (1, 10, 100 µg) of either V1Jns-wt-pol (virus derived) or V1R-wt-pol (codon optimized). At 5 wks post dose 1, 5 of 10 animals were randomly boosted with the same vaccine and dose they received initially. Figure 9 shows the geometric mean titers of the BALB/c cohorts determined at 2 wks past boost. No significant anti-RT titers can be observed from animals immunized with one or two doses of the wt-pol plasmid (virus derived). In contrast, animals vaccinated with the humanized gene construct gave cohort anti-RT titers (>1000) significantly above background levels at doses above 10 ug. The responses seen at 10 and 100 ug dose of V1R-wt-pol (codon optimized) were boosted approximately 10-fold with a second immunization, reaching titers as high as 10<sup>6</sup>.

Spleens from all mice in each of the cohorts were collected to be analyzed for IFN-γ secretion following stimulation with mixtures of either CD4+ peptide epitopes or CD8+ peptide epitopes. The results are shown in Figure 10. All wt-pol vaccinees did

not show any significant cellular response above the background controls. In contrast, strong antigen-stimulated IFN- $\gamma$  secretion were observed in a dose-responsive manner from animals vaccinated with one or two doses of 10 or more  $\mu g$  of the wt-pol (codon optimized) construct.

The present invention is not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the invention in addition to those described herein will become apparent to those skilled in the art from the foregoing description. Such modifications are intended to fall within the scope of the appended claims.

10

#### WHAT IS CLAIMED IS:

1. A pharmaceutically acceptable DNA vaccine composition, which comprises:

- (a) a DNA expression vector; and,
- 5 (b) a DNA molecule containing a codon optimized open reading frame encoding a Pol protein or inactivated Pol derivative thereof, wherein upon administration of the DNA vaccine to a host the Pol protein or inactivated Pol derivative is expressed and generates a cellular immune response against HIV-1 infection.

10

- 2. The DNA vaccine of claim 1 wherein the DNA molecule encodes wild type Pol.
- 3. The DNA vaccine of claim 2 wherein the DNA molecule comprises the nucleotide sequence as set forth in SEQ ID NO:1.
  - 4. The DNA vaccine of claim 3 which is V1Jns-wt-pol.
- 5. The DNA vaccine of claim 1 wherein the DNA molecule encodes an inactivated Pol derivative which contains a nucleotide sequence encoding a human tissue plasminogen activator leader peptide.
  - 6. The DNA vaccine of claim 5 wherein the DNA molecule comprises the nucleotide sequence as set forth in SEQ ID NO:5

25

- 7. The DNA vaccine of claim 6 which is V1Jns-tPA-wt-pol.
- 8. The DNA vaccine of claim 1 wherein the inactivated Pol protein contains at least one amino acid modification within each region of the Pol protein responsible for reverse transcriptase activity, RNase H activity and integrase activity, such that the inactivated Pol protein shows no substantial reverse transcriptase activity, RNase H activity and integrase activity.

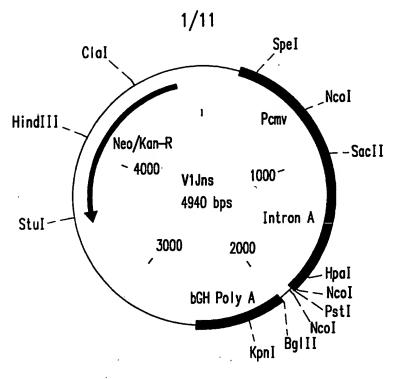
9. The DNA vaccine of claim 8 wherein the DNA molecule comprises the nucleotide sequence as set forth in SEQ ID NO:3

10. The DNA vaccine of claim 9 which is V1Jns-IAPol.

5

- 11. The DNA vaccine of claim 8 wherein the DNA molecule encodes an inactivated Pol derivative which contains a nucleotide sequence encoding a human tissue plasminogen activator leader peptide.
- 10 12. The DNA vaccine of claim 11 wherein the DNA molecule comprises the nucleotide sequence as set forth in SEQ ID NO:7.
  - 13. The DNA vaccine of claim 7 which is V1Jns-tPA-IAPol.
- 15 14. A method for inducing an immune response against infection or disease caused by virulent strains of HIV which comprises administering into the tissue of a mammalian host a pharmaceutically acceptable DNA vaccine composition which comprises a DNA expression vector and a DNA molecule containing a codon optimized open reading frame encoding a Pol protein or inactivated Pol derivative thereof, wherein upon administration of the DNA vaccine to the vertebrate host the Pol protein or inactivated Pol derivative is expressed and generates the immune response.
  - 15. The method of claim 16 wherein the mammalian host is a human.

- 16. The method of claim 17 wherein the DNA vaccine is selected from the group consisting of V1Jns-WTPol, V1Jns-tPA-WTPol, V1Jns-IAPol and V1Jns-tPA-IAPol.
- 30 17. A substantially purified protein which comprises an amino acid sequence selected from the group consisting of SEQ ID NO:4, SEQ ID NO:6, and SEQ ID NO:8.





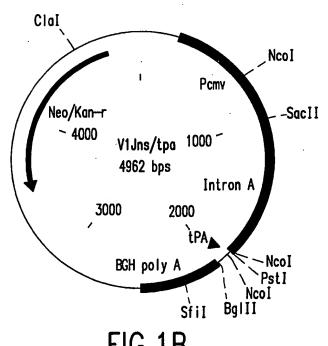


FIG.1B

SUBSTITUTE SHEET (RULE 26)

# 2/11

GCAGTGGCCCCTGACTGAGGAGAAGATCAAGGCCCTGGTGGAAATCTGCACTGAGATGGAGAAGGAGGGCAAAATCTCCA sGinTrpProLeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerL 30 40 50

AGATTGGCCCCGAGAACCCCTACAACACCCCTGTGTTTGCCATCAAGAAGAAGAAGACTCCACCAAGTGGAGGAAGCTGGTG
yslieGiyProGiuAsnProTyrAsnThrProVaiPheAialieLysLysLysAspSerThrLysTrpArgLysLeuVai
60 70

GACTTCAGGGAGCTGAACAAGAGGACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCTGGCCTGAAGAA AspPheArgGIuLeuAsnLysArgThrGInAspPheTrpGIuVaIGInLeuGIyIIeProHisProAloGIyLeuLysLy 80 90 100

GAAGAAGTCTGTGACTGTGCTGCTGTGGGGGATGCCTACTTCTCTGTGCCCCTGGATGAGGACTTCAGGAAGTACACTG sLysLysSerVaIThrVaILeu<u>AIa</u>VaIGIyAspAIaTyrPheSerVaIProLeuAspGIuAspPheArgLysTyrThrA 110 120 130

CCTTCACCATCCCCTCCATCAACAATGAGACCCCTGGCATCAGGTACCAGTACAATGTGCTGCCCCAGGGCTGGAAGGGC
InPheThrlieProSerileAsnAsnGluThrProGlylleArgTyrGInTyrAsnValLeuProGlnGlyTrpLysGly
140 150

TCCCCTGCCATCTTCCAGTCCTCCATGACCAAGATCCTGGAGCCCTTCAGGAAGCAGAACCCTGACATTGTGATCTACCA SerProAloIIePheGInSerSerMetThrLysIIeLeuGIuProPheArgLysGInAsnProAspIIeVaiIieTyrGI 160 170 180

GTACATGGCTGCCCTGTATGTGGGCTCTGACCTGGAGATTGGGCAGCACAGGACCAAGATTGAGGAGCTGAGGCAGCACCC
nTyrMetAlaAlaLeuTyrVaIGIySerAspLeuGIuIIeGIyGInHisArgThrLysIIeGIuGIuLeuArgGInHisL
190 200 210

TGCTGAGGTGGGGCCTGACCACCCCTGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTATGAGCTGCAC euLeuArgTrpG1yLeuThrThrProAspLysLysHisG1nLysG1uProProPheLeuTrpMetG1yTyrG1uLeuHis 220 230

CCCGACAAGTGGACTGTGCAGCCCATTGTGCTGCCTGAGAAGGACTCCTGGACTGTGAATGACATCCAGAAGCTGGTGGG ProAspLysTrpThrVaIGInProIleVaILeuProGIuLysAspSerTrpThrVaIAsnAspIieGInLysLeuVaIGI 240 250 260

CAAGCTGAACTGGGCCTCCCAAATCTACCCTGGCATCAAGGTGAGGCAGCTGTGCAAGCTGCTGAGGGGCACCAAGGCCC
yLysLeuAsnTrpAIaSerGInIIeTyrProGIyIIeLysVaIArgGInLeuCysLysLeuLeuArgGIyThrLysAIaL
270 280 290

FIG.2A

SUBSTITUTE SHEET (RULE 26)

TGACTGAGGTGATCCCCCTGACTGAGGAGGCTGAGCTGGAGCTGGGCTGAGAACAGGGAGATCCTGAAGGAGCCTGTGCAT EuThrGluVallleProLeuThrGluGluAlaGluLeuGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHis 300 310

GCGGTGTACTATGACCCCTCCAAGGACCTGATTGCTGAGATCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAAATCTA GlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTy 320 330 340

CCAGGAGCCCTTCAAGAACCTGAAGACTGGCAAGTATGCCAGGATGACGGGGGCCCCACACCAATGATGTGAAGCAGCTGA rGInGIuProPheLysAsnLeuLysThrGIyLysTyrAIaArgMetArgGIyAIaHisThrAsnAspVaILysGInLeuT 350 360 370

CTGAGGCTGTGCAGAAGATCACCACTGAGTCCATTGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAG hrGIuAlaVaIGInLysIIeThrThrGIuSerIIeVaIIIeTrpGIyLysThrProLysPheLysLeuProIIeGInLys 380 390

GCTGAAGCTGTGCTACCAGCTGGAGAAGGAGCCCATTGTGGGGGCCTGAGACCTTCTATGTGGCTGGGGCTGCCAACAGGG uValLysLeuTrpTyrGInLeuGIuLysGIuProIIeVaIGIyAlaGIuThrPheTyrVaIAIaGIyAlaAlaAsnArgG 430 440 450

AAGACTGCCCTCCAGGCCATCTACCTGGCCCTCCAGGACTCTGGCCTGGAGGTGAACATTGTGACTGCCTCCCAGTATGC
LysThrAigLeuGinAlaileTyrLeuAlaLeuGinAspSerGlyLeuGiuValAsnIleValThrAigSerGInTyrAl
480
490
500

CCTGGGCATCATCCAGGCCCAGCCTGATCAGTCTGAGTCTGAGCTGGTGAACCAGATCATTGAGCAGCTGATCAAGAAGG aLeuGlyIleIleGlnAloGlnProAspGInSerGluSerGluLeuValAsnGInIleIleGluGInLeuIleLysLysG 510 520 530

AGAAGGTGTACCTGGCCTGCCCACAAGGGCATTGGGGGCAATGAGCAGGTGGACAAGCTGGTGTCTGCTGGC
IuLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGly
540
550

ATCAGGAAGGTGCTGTTCCTGGATGGCATTGACAAGGCCCAGGATGAGCATGAGAAGTACCACTCCAACTGGAGGGCTAT ITeArgLysVaILeuPheLeuAspG1yI1eAspLysA1aG1nAspG1uHisG1uLysTyrHisSerAsnTrpArgA1aMe 560 570 580

FIG.2B

**SUBSTITUTE SHEET (RULE 26)** 

CGCCTCTGACTTCAACCTGCCCCCTGTGGTGGCTAAGGAGATTGTGGCCTCCTGTGACAAGTGCCAGCTGAAGGGGGAGG tAlaSerAspPheAsnLeuProProValValAlaLysGlulleValAlaSerCysAspLysCysGlnLeuLysGlyGluA 590 600 610

GCTGTGCATGTGGCCTCCGGCTACATTGAGGCTGAGGTGATCCCTGCTGAGACAGGCCAGGAGACTGCCTACTTCCTGCT AlaValHisValAlaSerGlyTyr[leGluAlaGluVallleProAlaGluThrGlyGlnGluThrAlaTyrPheLeuLe 640 650 660

GAAGCTGGCTGCCAGGTGCCCTGTGAAGACCATCCAEACTGCCAATGGCTCCAACTTCACTGGGGCCACAGTGAGGGCTG
uLysLeuAloGlyArgTrpProVaiLysThrlleHisThrAlaAsnGlySerAsnPheThrGlyAlaThrValArgAlaA
670 680 690

CCTGCTGGTGGCCTCCAGGAGTTTGGCATCCCCTACAACCCCCAGTCCCAGGGGGTGGTGGCCTCCATGAAC laCysTrpTrpAlaGlyIleLysGlnGluPheGlyIleProTyrAsnProGlnSerGlnGlyValValAlaSerMetAsn 700 710

AAGGAGCTGAAGAAGATCATTGGGCAGGTGAGGGACCAGGCTGAGCACCTGAAGACAGCTGTGCAGATGGCTGTGTTCAT LysGIuLeuLysLysIIeIIeGIyGInVaIArgAspGInAIaGIuHisLeuLysThrAIaVaIGInMetAIaVaIPheII 720 730 740

CCACAACTTCAAGAGGAAGGGGGGCATCGGGGGCTACTCCGCTGGGGAGAGGATTGTGGACATCATTGCCACAGACATCC
eHisAsnPheLysArgLysGlyGlyIleGlyGlyTyrSerAlaGlyGluArgIleValAspIleIleAlaThrAspIleG
750 760 770

AGACCAAGGAGCTCCAGAAGCAGATCACCAAGATCCAGAACTTCAGGGTGTACTACAGGGACTCCAGGAACCCCCTGTGG
InThrLysGluLeuGInLysGInIleThrLysIleGInAsnPheArgValTyrTyrArgAspSerArgAsnProLeuTrp
780 790

AAAGCCCGGGCAGATCT (SEQ ID NO: 3) Xx = BgIII

FIG.2C

**SUBSTITUTE SHEET (RULE 26)** 

GATCACCATGGATGCAATGAAGAGGGCCTCTGCTGTGCTGCTGTGTGGAGCAGTCTTCGTTTCGC MetAspAlaMetLysArgGlyLeuCysCysValLeuLeuCysGlyAlaValPheValSerP -25 RoSerGIuIleSerAlaProIleSerProIleGIuThrVaIProVaILysLeuLysProGlyMetAspGly 20 20

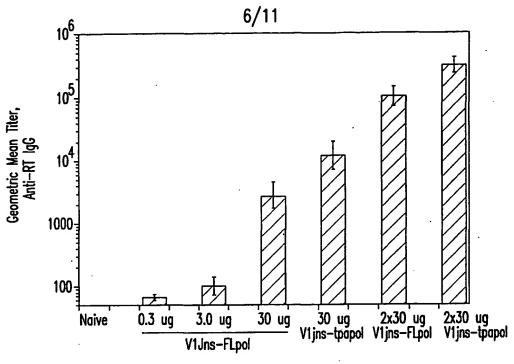


FIG.4

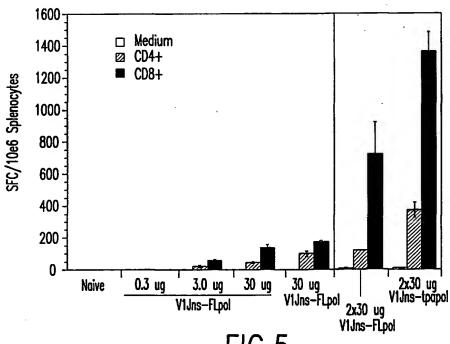
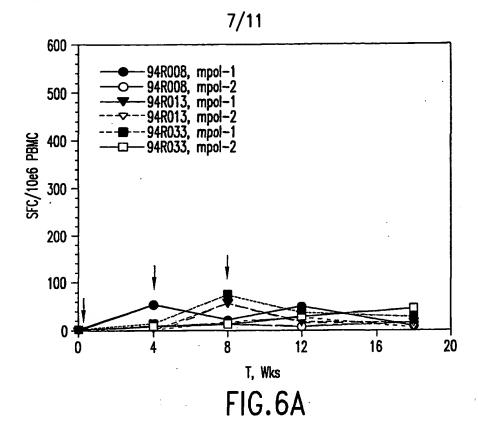


FIG.5
SUBSTITUTE SHEET (RULE 26)



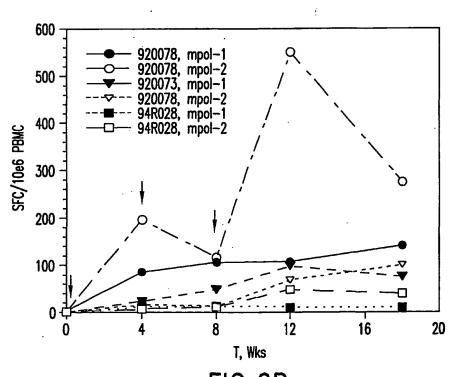
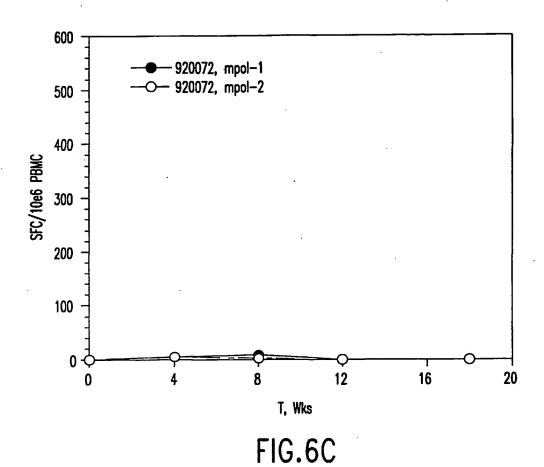
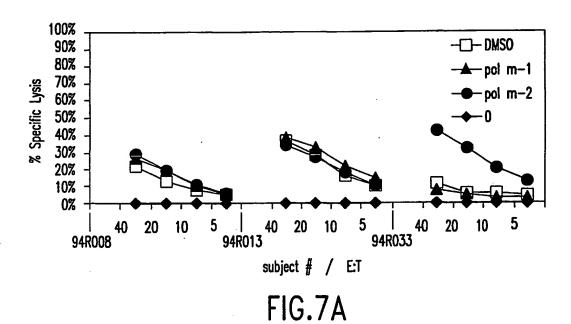
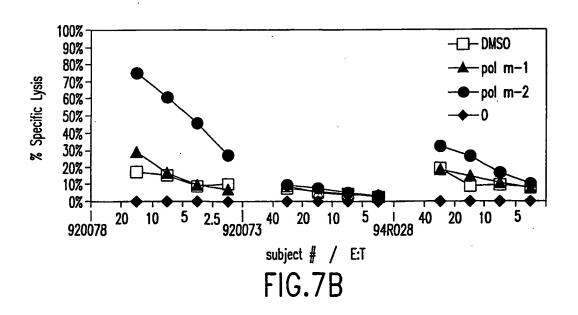


FIG.6B SUBSTITUTE SHEET (RULE 26)



SUBSTITUTE SHEET (RULE 26)





**SUBSTITUTE SHEET (RULE 26)** 

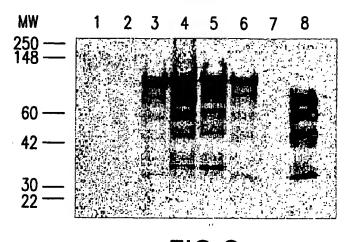
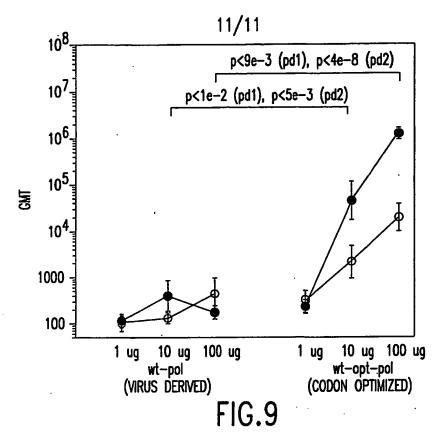


FIG.8



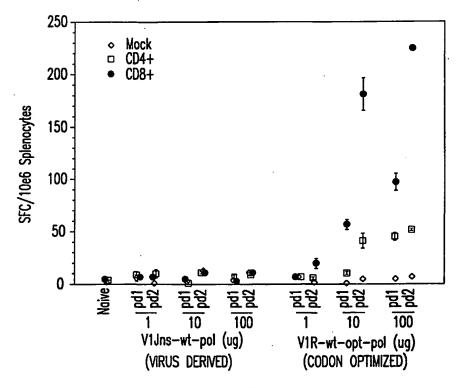


FIG. 10 SUBSTITUTE SHEET (RULE 26)

## SEQUENCE LISTING

<110> Merck & Co., Inc.

130

<120> POLYNUCLEOTIDE VACCINES EXPRESSING CODON OPTIMIZED HIV-1 POL AND MODIFIED HIV-1 POL

<130> 20608Y PCT <160> 30 <170> FastSEQ for Windows Version 4.0 <211> 2577 <212> DNA <213> Human Immunodeficiency Virus-1 <221> CDS <222> (10)...(2562) <400> 1 agatctacc atg gcc ccc atc tcc ccc att gag act gtg cct gtg aag ctg Met Ala Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu  $1 \hspace{1cm} 5 \hspace{1cm} 10$ aag cet gge atg gat gge eec aag gtg aag eag tgg eec etg act gag Lys Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu 99 gag aag atc aag gcc ctg gtg gaa atc tgc act gag atg gag aag gag Glu Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu 147 ggc aaa atc tcc aag att ggc ccc gag aac ccc tac aac acc cct gtg Gly Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val 195 ttt gcc atc aag aag gac tcc acc aag tgg agg aag ctg gtg gac 243 Phe Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp ttc agg gag ctg aac aag agg acc cag gac ttc tgg gag gtg cag ctg Phe Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu 291 ggc atc ccc cac ccc gct ggc ctg aag aag aag tct gtg act gtg Gly Ile Pro His Pro Ala Gly Leu Lys Lys Lys Ser Val Thr Val 339 ctg gat gtg ggg gat gcc tac ttc tct gtg ccc ctg gat gag gac ttc Leu Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe 387 agg aag tac act gcc ttc acc atc ccc tcc atc aac aat gag acc cct 435 Arg Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro

135

ggc Gly	atc Ile	agg Arg 145	tac Tyr	cag Gln	tac Tyr	aat Asn	gtg Val 150	ctg Leu	ccc Pro	cag Gln	ggc Gly	tgg Trp 155	aag Lys	ggc Gly	tcc Ser	483
					tcc Ser											531
					att Ile 180											579
					gag Glu											627
					ctg Leu											675
					ccc Pro											723
					cag Gln										tgg Trp	771
					cag Gln 260											819
					atc Ile											867
					act Thr											915
					aac Asn											963
					tcc Ser											1011
					acc Thr 340											1059
					tat Tyr											1107
					gag Glu											1155

					acc Thr											1203
					tgg Trp											1251
gag Glu 415	tgg Trp	gag Glu	ttt Phe	gtg Val	aac Asn 420	acc Thr	ccc Pro	ccc Pro	ctg Leu	gtg Val 425	aag Lys	ctg Leu	tgg Trp	tac Tyr	cag Gln 430	1299
					att Ile											1347
					acc Thr											1395
					gtg Val											1443
act Thr	gag Glu 480	ctc Leu	cag Gln	gcc Ala	atc Ile	tac Tyr 485	ctg Leu	gcc Ala	ctc Leu	cag Gln	gac Asp 490	tct Ser	ggc Gly	ctg Leu	gag Glu	1491
					gac Asp 500											1539
					gag Glu			Leu								1587
					aag Lys											1635
					gag Glu											1683
					ctg Leu											1731
					aac Asn 580											1779
ccc Pro	cct Pro	gtg Val	gtg Val	gct Ala 595	aag Lys	gag Glu	att Ile	gtg Val	gcc Ala 600	tcc Ser	tgt Cys	gac Asp	aag Lys	tgc Cys 605	cag Gln	1827
					atg Met											1875

			gac Asp													1923
gtg Val	cat His 640	gtg Val	gcc Ala	tcc Ser	Gly ggc	tac Tyr 645	att Ile	gag Glu	gct Ala	gag Glu	gtg Val 650	atc Ile	cct Pro	gct Ala	gag Glu	1971
aca Thr 655	ggc Gly	cag Gln	gag Glu	act Thr	gcc Ala 660	tac Tyr	ttc Phe	ctg Leu	ctg Leu	aag Lys 665	ctg Leu	gct Ala	ggc Gly	agg Arg	tgg Trp 670	2019
cct Pro	gtg Val	aag Lys	acc Thr	atc Ile 675	cac His	act Thr	gac Asp	aat Asn	ggc Gly 680	tcc Ser	aac Asn	ttc Phe	act Thr	ggg Gly 685	gcc Ala	2067
			gct Ala 690													2115
atc Ile	ccc Pro	tac Tyr 705	aac Asn	ccc Pro	cag Gln	tcc Ser	cag Gln 710	Gly ggg	gtg Val	gtg Val	gag Glu	tcc Ser 715	atg Met	aac Asn	aag Lys	2163
			aag Lys													2211.
			gtg Val													2259
Gly	gly	atc Ile	Gly ggg	ggc Gly 755	tac Tyr	tcc Ser	gct Ala	Gly ggg	gag Glu 760	agg Arg	att Ile	gtg Val	gac Asp	atc Ile 765	att Ile	2307
			atc Ile 770													2355
cag Gln	aac Asn	ttc Phe 785	agg Arg	gtg Val	tac Tyr	tac Tyr	agg Arg 790	gac Asp	tcc Ser	agg Arg	aac Asn	ccc Pro 795	ctg Leu	tgg Trp	aag Lys	2403
			aag Lys													2451
			gac Asp													2499
			ggc Gly													2547
	gat Asp		gac Asp 850	taa *	agco	cggg	gca g	gatct	:							2577

<210> 2

<211> 850

```
<212> PRT
<213> Human Immunodeficiency Virus-1
Met Ala Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys Pro
                                       10
Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu Lys
             20
Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly Lys
       35
                              40
Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe Ala 50 55
Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg 65 70 75 80
Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile
85 90 95
Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu Asp
100 105 110
Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg Lys
115 120 125
Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly Ile
130 135 140
Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala
145 150 155 160
Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys Gln
165 170 175
Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val Gly
180 185 190
                                  185
Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu Arg
        195
                             200
                                                   205
Gln His Leu Leu Arg Trp Gly Leu Thr Thr Pro Asp Lys Lys His Gln 210 220
Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys 225 230 235 240
Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr Val 245 250 255
Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile
           260
                                  265
Tyr Pro Gly Ile Lys Val Arg Gln Leu Cys Lys Leu Leu Arg Gly Thr 275 280 285
                                                   285
Lys Ala Leu Thr Glu Val Ile Pro Leu Thr Glu Glu Ala Glu Leu Glu
    290
                          295
Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro Val His Gly Val Tyr
                    310
                                           315
                                                                 320
Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile Gln Lys Gln Gly Gln
                325
                                      330
                                                            335
Gly Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu Lys 340 345 350
                                  345
                                                        350
Thr Gly Lys Tyr Ala Arg Met Arg Gly Ala His Thr Asn Asp Val Lys 355 360 365
Gln Leu Thr Glu Ala Val Gln Lys Ile Thr Thr Glu Ser Ile Val Ile
    370
                         375
                                               380
Trp Gly Lys Thr Pro Lys Phe Lys Leu Pro Ile Gln Lys Glu Thr Trp 385 390 395 400
Glu Thr Trp Trp Thr Glu Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp
                405
                                      410
                                                            415
Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu Glu
            420
                                  425
                                                       430
Lys Glu Pro Ile Val Gly Ala Glu Thr Phe Tyr Val Asp Gly Ala Ala
        435
                              440
```

```
Asn Arg Glu Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Asn Arg Gly
    450
                         455
                                                 460
Arg Gln Lys Val Val Thr Leu Thr Asp Thr Thr Asn Gln Lys Thr Glu
                   470
                                           475
                                                                    480
465
Leu Gln Ala Ile Tyr Leu Ala Leu Gln Asp Ser Gly Leu Glu Val Asn
485 490 495
                485
Ile Val Thr Asp Ser Gln Tyr Ala Leu Gly Ile Ile Gln Ala Gln Pro
                                   505
            500
                                                          510
Asp Gln Ser Glu Ser Glu Leu Val Asn Gln Ile Ile Glu Gln Leu Ile
515 520 525
Lys Lys Glu Lys Val Tyr Leu Ala Trp Val Pro Ala His Lys Gly Ile
                          535
                                              540
Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser Ala Gly Ile Arg Lys 545 555 560
                                                                    560
Val Leu Phe Leu Asp Gly Ile Asp Lys Ala Gln Asp Glu His Glu Lys
565 570 575
Tyr His Ser Asn Trp Arg Ala Met Ala Ser Asp Phe Asn Leu Pro Pro 580 590
Val Val Ala Lys Glu Ile Val Ala Ser Cys Asp Lys Cys Gln Leu Lys
595 600 605
Gly Glu Ala Met His Gly Gln Val Asp Cys Ser Pro Gly Ile Trp Gln 610 620
                         615
Leu Asp Cys Thr His Leu Glu Gly Lys Val Ile Leu Val Ala Val His 625 630 635 640
                  630
                                          635
Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile Pro Ala Glu Thr Gly
645 650 655
Gln Glu Thr Ala Tyr Phe Leu Leu Lys Leu Ala Gly Arg Trp Pro Val660 665 670
Lys Thr Ile His Thr Asp Asn Gly Ser Asn Phe Thr Gly Ala Thr Val
675 680 685
Arg Ala Ala Cys Trp Trp Ala Gly Ile Lys Gln Glu Phe Gly Ile Pro 690 700
Tyr Asn Pro Gln Ser Gln Gly Val Val Glu Ser Met Asn Lys Glu Leu 705 710 720
Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala Glu His Leu Lys Thr 725 \hspace{1.5cm} 730 \hspace{1.5cm} 735
Ala Val Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys Gly Gly 740 745 750
Ile Gly Gly Tyr Ser Ala Gly Glu Arg Ile Val Asp Ile Ile Ala Thr 755 760 765
Asp Ile Gln Thr Lys Glu Leu Gln Lys Gln Ile Thr Lys Ile Gln Asn 770 780
Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asn Pro Leu Trp Lys Gly Pro 785 790 800
                     790
                                           795
Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val Val Ile Gln Asp Asn 805 810
Ser Asp Ile Lys Val Val Pro Arg Arg Lys Ala Lys Ile Ile Arg Asp 820 825 830
Tyr Gly Lys Gln Met Ala Gly Asp Asp Cys Val Ala Ser Arg Gln Asp
                               840
Glu Asp
  850
<210> 3
<211> 2577
<212> DNA
<213> Human Immunodeficiency Virus-1
<220>
<221> CDS
<222> (10)...(2562)
```

	0> 3 tctad	Me									ar Va				ag ctg ys Leu	
	cct Pro															99
	aag Lys															147
	aaa Lys															195
	gcc Ala															243
	agg Arg 80														ctg Leu	291
ggc Gly 95	atc Ile	ccc Pro	cac His	ccc Pro	gct Ala 100	ggc Gly	ctg Leu	aag Lys	aag Lys	aag Lys 105	aag Lys	tct Ser	gtg Val	act Thr	gtg Val 110	339
	gct Ala															387
	aag Lys															435
	atc Ile															483
	gcc Ala 160															531
aag Lys 175	cag Gln	aac Asn	cct Pro	gac Asp	att Ile 180	gtg Val	atc Ile	tac Tyr	cag Gln	tac Tyr 185	atg Met	gct Ala	gcc Ala	ctg Leu	tat Tyr 190	579
	ggc Gly															627
ctg Leu	agg Arg	cag Gln	cac His 210	ctg Leu	ctg Leu	agg Arg	tgg Trp	ggc Gly 215	ctg Leu	acc Thr	acc Thr	cct Pro	gac Asp 220	aag Lys	aag Lys	675
cac His	cag Gln	aag Lys 225	gag Glu	ccc Pro	ccc Pro	ttc Phe	ctg Leu 230	tgg Trp	atg Met	ggc Gly	tat Tyr	gag Glu 235	ctg Leu	cac His	ccc Pro	723

gac a Asp I	aag Lys 240	tgg Trp	act Thr	gtg Val	cag Gln	ccc Pro 245	att Ile	gtg Val	ctg Leu	cct Pro	gag Glu 250	aag Lys	gac Asp	tcc Ser	tgg Trp	771
act of Thr N 255	gtg Val	aat Asn	gac Asp	atc Ile	cag Gln 260	aag Lys	ctg Leu	gtg Val	ggc Gly	aag Lys 265	ctg Leu	aac Asn	tgg Trp	gcc Ala	tcc Ser 270	819
caa a Gln l	atc Ile	tac Tyr	cct Pro	ggc Gly 275	atc Ile	aag Lys	gtg Val	agg Arg	cag Gln 280	ctg Leu	tgc Cys	aag Lys	ctg Leu	ctg Leu 285	agg Arg	867
ggc a																915
ctg g Leu (																963
gtg t Val 1																1011
ggc o Gly o 335						Tyr										1059
ctg a Leu I																1107
gtg a																1155
gtg a Val 1																1203
acc t Thr T																1251
gag t Glu T 415																1299
ctg g Leu G																1347
gct g Ala A																1395
agg g Arg G																1443

			cag Gln														1491
			gtg Val														1539
cag Gln	cct Pro	gat Asp	cag Gln	tct Ser 515	gag Glu	tct Ser	gag Glu	ctg Leu	gtg Val 520	aac Asn	cag Gln	atc Ile	att Ile	gag Glu 525	cag Gln		1587
			aag Lys 530												aag Lys	-	1635
ggc Gly	att Ile	ggg Gly 545	ggc Gly	aat Asn	gag Glu	cag Gln	gtg Val 550	gac Asp	aag Lys	ctg Leu	gtg Val	tct Ser 555	gct Ala	ggc Gly	atc Ile		1683
			ctg Leu														1731
gag Glu 575	aag Lys	tac Tyr	cac His	tcc Ser	aac Asn 580	tgg Trp	agg Arg	gct Ala	atg Met	gcc Ala 585	tct Ser	gac Asp	ttc Phe	aac Asn	ctg Leu 590		1779
			gtg Val														·1827
			gag Glu 610												atc Ile	•	1875
			gcc Ala														1923
			gcc Ala														1971
			gag Glu														2019
cct Pro	gtg Val	aag Lys	acc Thr	atc Ile 675	cac His	act Thr	gcc Ala	aat Asn	ggc 680	tcc Ser	aac Asn	ttc Phe	act Thr	ggg Gly 685	gcc Ala		2067
			gct Ala 690														2115
atc Ile	ccc Pro	tac Tyr 705	aac Asn	ccc Pro	cag Gln	tcc Ser	cag Gln 710	Gly ggg	gtg Val	gtg Val	gcc Ala	tcc Ser 715	atg Met	aac Asn	aag Lys		2163

gag Glu	ctg Leu 720	aag Lys	aag Lys	atc Ile	att Ile	ggg Gly 725	cag Gln	gtg Val	agg Arg	gac Asp	cag Gln 730	gct Ala	gag Glu	cac His	ctg Leu	2211
					atg Met 740											2259
Gly	ggc	atc Ile	Gly ggg	ggc Gly 755	tac Tyr	tcc Ser	gct Ala	Gly ggg	gag Glu 760	agg Arg	att Ile	gtg Val	gac Asp	atc Ile 765	att Ile	2307
					acc Thr											2355
cag Gln	aac Asn	ttc Phe 785	agg Arg	gtg Val	tac Tyr	tac Tyr	agg Arg 790	gac Asp	tcc Ser	agg Arg	aac Asn	ccc Pro 795	ctg Leu	tgg Trp	aag Lys	2403
ggc Gly	cct Pro 800	gcc Ala	aag Lys	ctg Leu	ctg Leu	tgg Trp 805	aag Lys	Gly ggg	gag Glu	ggg Gly	gct Ala 810	gtg Val	gtg Val	atc Ile	cag Gln	2451
					aag Lys 820											2499
					cag Gln											2547
	_	gag Glu	-	taa *	agco		gca g	gatct	:							2577
<21	L> 85 2> PI	RT	Imm	4.												
<400				mode	efici	ency	/ Vii	rus-1	Ĺ							
Met	)> 4 Ala	Pro			efici Pro				Va1	Pro	Val	Lys	Leu		Pro	٠
Met 1	Ala		Ile Gly	Ser 5		Ile	Glu	Thr Gln	Val 10			=	Glu	15		
Met 1 Gly	Ala Met	Asp Ala	Ile Gly 20	Ser 5 Pro	Pro	Ile Val	Glu Lys Cys	Thr Gln 25	Val 10 Trp	Pro	Leu	Thr Lys	Glu 30	15 Glu	Lys	
Met 1 Gly Ile	Ala Met Lys Ser	Asp Ala 35	Ile Gly 20 Leu	Ser 5 Pro Val	Pro Lys	Ile Val Ile Glu	Glu Lys Cys 40	Thr Gln 25 Thr	Val 10 Trp Glu	Pro Met	Leu Glu Thr	Thr Lys 45	Glu 30 Glu	15 Glu Gly	Lys Lys	
Met 1 Gly Ile Ile	Ala Met Lys Ser 50	Asp Ala 35 Lys	Ile Gly 20 Leu Ile	Ser 5 Pro Val Gly	Pro Lys Glu	Ile Val Ile Glu 55	Glu Lys Cys 40 Asn	Thr Gln 25 Thr	Val 10 Trp Glu Tyr	Pro Met Asn	Leu Glu Thr 60	Thr Lys 45 Pro	Glu 30 Glu Val	15 Glu Gly Phe	Lys Lys Ala	
Met 1 Gly Ile Ile 65	Ala Met Lys Ser 50 Lys	Asp Ala 35 Lys Lys	Ile Gly 20 Leu Ile Lys	Ser 5 Pro Val Gly Asp	Pro Lys Glu Pro Ser	Ile Val Ile Glu 55 Thr	Glu Lys Cys 40 Asn Lys	Thr Gln 25 Thr Pro	Val 10 Trp Glu Tyr Arg	Pro Met Asn Lys 75	Leu Glu Thr 60 Leu	Thr Lys 45 Pro Val	Glu 30 Glu Val Asp	15 Glu Gly Phe	Lys Lys Ala Arg 80	
Met 1 Gly Ile Ile 65 Glu	Ala Met Lys Ser 50 Lys Leu	Asp Ala 35 Lys Lys Asn	Ile Gly 20 Leu Ile Lys Lys Ala	Ser 5 Pro Val Gly Asp Arg 85	Pro Lys Glu Pro Ser 70	Ile Val Ile Glu 55 Thr	Glu Lys Cys 40 Asn Lys Asp	Thr Gln 25 Thr Pro Trp Phe Lys	Val 10 Trp Glu Tyr Arg	Pro Met Asn Lys 75 Glu	Leu Glu Thr 60 Leu Val	Thr Lys 45 Pro Val Gln	Glu 30 Glu Val Asp Leu	15 Glu Gly Phe Phe Gly 95	Lys Lys Ala Arg 80 Ile	
Met 1 Gly Ile Ile 65 Glu Pro	Ala Met Lys Ser 50 Lys Leu His	Asp Ala 35 Lys Lys Asn Pro	Ile Gly 20 Leu Ile Lys Lys Ala 100	Ser 5 Pro Val Gly Asp Arg 85 Gly	Pro Lys Glu Pro Ser 70 Thr	Ile Val Ile Glu 55 Thr Gln Lys	Glu Lys Cys 40 Asn Lys Asp	Thr Gln 25 Thr Pro Trp Phe Lys 105	Val 10 Trp Glu Tyr Arg Trp 90 Lys	Pro Met Asn Lys 75 Glu Ser	Leu Glu Thr 60 Leu Val	Thr Lys 45 Pro Val Gln	Glu 30 Glu Val Asp Leu Val 110	15 Glu Gly Phe Phe Gly 95 Leu	Lys Lys Ala Arg 80 Ile Ala	

Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala 145 150 150 160 150 155 Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys Gln 165 170 175 165 Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Ala Ala Leu Tyr Val Gly 180 185 190 Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu Arg 195 200 205 Ser Asp Leu Giu lie Gly Gin 200 205

Gln His Leu Leu Arg Trp Gly Leu Thr Thr Pro Asp Lys His Gln 215 220 Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys 225 230 235 240 Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr Val 245 250 255 Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile 260 265 270 Tyr Pro Gly Ile Lys Val Arg Gln Leu Cys Lys Leu Leu Arg Gly Thr 275 280 285 Lys Ala Leu Thr Glu Val Ile Pro Leu Thr Glu Glu Ala Glu Leu Glu 290 295 300 Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro Val His Gly Val Tyr 310 315 320 Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile Gln Lys Gln Gly Gln 325 330 Gly Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu Lys 340 345 350350 Thr Gly Lys Tyr Ala Arg Met Arg Gly Ala His Thr Asn Asp Val Lys 355 360 365 Gln Leu Thr Glu Ala Val Gln Lys Ile Thr Thr Glu Ser Ile Val Ile 370 380 Trp Gly Lys Thr Pro Lys Phe Lys Leu Pro Ile Gin Lys Glu Thr Trp 385 390 395 400 Glu Thr Trp Trp Thr Glu Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp 405 410 Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu Glu
420 425 430 420 425 Lys Glu Pro Ile Val Gly Ala Glu Thr Phe Tyr Val Ala Gly Ala Ala 445 Asn Arg Glu Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Asn Arg Gly 450 460 460 Arg Gln Lys Val Val Thr Leu Thr Asp Thr Thr Asn Gln Lys Thr Ala 465 470 475 480 Leu Gln Ala Ile Tyr Leu Ala Leu Gln Asp Ser Gly Leu Glu Val Asn 485 490 495 Ile Val Thr Ala Ser Gln Tyr Ala Leu Gly Ile Ile Gln Ala Gln Pro
500 505 505 Asp Gln Ser Glu Ser Glu Leu Val Asn Gln Ile Ile Glu Gln Leu Ile 515 520 525 Lys Lys Glu Lys Val Tyr Leu Ala Trp Val Pro Ala His Lys Gly Ile 530 540 Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser Ala Gly Ile Arg Lys 545 550 560 Val Leu Phe Leu Asp Gly Ile Asp Lys Ala Gln Asp Glu His Glu Lys 565 570 575 Tyr His Ser Asn Trp Arg Ala Met Ala Ser Asp Phe Asn Leu Pro Pro 580 585 590 Val Val Ala Lys Glu Ile Val Ala Ser Cys Asp Lys Cys Gln Leu Lys 595 600 605 Gly Glu Ala Met His Gly Gln Val Asp Cys Ser Pro Gly Ile Trp Gln 610 620 Leu Ala Cys Thr His Leu Glu Gly Lys Val Ile Leu Val Ala Val His 635

```
Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile Pro Ala Glu Thr Gly
                  645
                                         650
                                                                  655
Gln Glu Thr Ala Tyr Phe Leu Leu Lys Leu Ala Gly Arg Trp Pro Val
                                                            670
             660
                                    665
Lys Thr Ile His Thr Ala Asn Gly Ser Asn Phe Thr Gly Ala Thr Val
                                680
Arg Ala Ala Cys Trp Trp Ala Gly Ile Lys Gln Glu Phe Gly Ile Pro
690 695 700
Tyr Asn Pro Gln Ser Gln Gly Val Val Ala Ser Met Asn Lys Glu Leu
705 710 715 720
                     710
                                             715 .
Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala Glu His Leu Lys Thr
725 730 735
Ala Val Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys Gly Gly 740 745 750
Ile Gly Gly Tyr Ser Ala Gly Glu Arg Ile Val Asp Ile Ile Ala Thr
755 760 765
Asp Ile Gln Thr Lys Glu Leu Gln Lys Gln Ile Thr Lys Ile Gln Asn 770 780
Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asn Pro Leu Trp Lys Gly Pro 785 790 795 800
Ala Lys Leu Trp Lys Gly Glu Gly Ala Val Val Ile Gln Asp Asn
                 805
                                         810
                                                                 815
Ser Asp Ile Lys Val Val Pro Arg Arg Lys Ala Lys Ile Ile Arg Asp
                                    825
Tyr Gly Lys Gln Met Ala Gly Asp Asp Cys Val Ala Ser Arg Gln Asp
        835
                                 840
Glu Asp
    850
<210> 5
<211> 2650
<212> ·DNA
<213> Human Immunodeficiency Virus-1
<220>
<221> CDS
<222> (8)...(2635)
<400> 5
gatcacc atg gat gca atg aag aga ggg ctc tgc tgt gtg ctg ctg ctg
Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu
                                                                                  49
tgt gga gca gtc ttc gtt tcg ccc agc gag atc tcc gcc ccc atc tcc
                                                                                  97
Cys Gly Ala Val Phe Val Ser Pro Ser Glu Ile Ser Ala Pro Ile Ser
                        20
ccc att gag act gtg cct gtg aag ctg aag cct ggc atg gat ggc ccc
Pro Ile Glu Thr Val Pro Val Lys Leu Lys Pro Gly Met Asp Gly Pro
                                                                                 145
aag gtg aag cag tgg ccc ctg act gag gag aag atc aag gcc ctg gtg
                                                                                 193
Lys Val Lys Gln Trp Pro Leu Thr Glu Glu Lys Ile Lys Ala Leu Val
gaa atc tgc act gag atg gag aag ggc aaa atc tcc aag att ggc Glu Ile Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser Lys Ile Gly
                                                                                 241
ccc gag aac ccc tac aac acc cct gtg ttt gcc atc aag aag aag gac
                                                                                 289
Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe Ala Ile Lys Lys Lys Asp
```

tcc Ser 95	acc Thr	aag Lys	tgg Trp	agg Arg	aag Lys 100	ctg Leu	gtg Val	gac Asp	ttc Phe	agg Arg 105	gag Glu	ctg Leu	aac Asn	aag Lys	agg Arg 110	337
acc Thr	cag Gln	gac Asp	ttc Phe	tgg Trp 115	gag Glu	gtg Val	cag Gln	ctg Leu	ggc Gly 120	atc Ile	ccc Pro	cac His	ccc Pro	gct Ala 125	ggc Gly	385
					tct Ser											. 433
ttc Phe	tct Ser	gtg Val 145	ccc Pro	ctg Leu	gat Asp	gag Glu	gac Asp 150	ttc Phe	agg Arg	aag Lys	tac Tyr	act Thr 155	gcc Ala	ttc Phe	acc Thr	481
atc Ile	ecc Pro 160	tcc Ser	atc Ile	aac Asn	aat Asn	gag Glu 165	acc Thr	cct Pro	ggc Gly	atc Ile	agg Arg 170	tac Tyr	cag Gln	tac Tyr	aat Asn	529
gtg Val 175	ctg Leu	ccc Pro	cag Gln	ggc Gly	tgg Trp 180	aag Lys	ggc	tcc Ser	cct Pro	gcc Ala 185	atc Ile	ttc Phe	cag Gln	tcc Ser	tcc Ser 190	577
					gag Glu											625
					gat Asp											673
					aag Lys											721
					cct Pro											769
ctg Leu 255	tgg Trp	atg Met	ggc Gly	tat Tyr	gag Glu 260	ctg Leu	cac His	ccc Pro	gac Asp	aag Lys 265	tgg Trp	act Thr	gtg Val	cag Gln	ccc Pro 270	817
					aag Lys											865
					aac Asn											913
					aag Lys											961
					gag Glu											1009

gag Glu 335	atc Ile	ctg Leu	aag Lys	gag Glu	cct Pro 340	gtg Val	cat His	Gly ggg	gtg Val	tac Tyr 345	tat Tyr	gac Asp	ccc Pro	tcc Ser	aag Lys 350		1057
										cag Gln							1105
										aag Lys							1153
agg Arg	atg Met	agg Arg 385	ggg	gcc Ala	cac His	acc Thr	aat Asn 390	gat Asp	gtg Val	aag Lys	cag Gln	ctg Leu 395	act Thr	gag Glu	gct Ala		1201
Val										atc Ile							1249
										tgg Trp 425							1297
										tgg Trp							1345
		Leu								gag Glu							1393
										gcc Ala							1441
ctg Leu	ggc Gly 480	aag Lys	gct Ala	ggc Gly	tat Tyr	gtg Val 485	acc Thr	aac Asn	agg Arg	ggc Gly	agg Arg 490	cag Gln	aag Lys	gtg Val	gtg Val		1489
										gag Glu 505							1537
										aac Asn							1585
										cct Pro							1633
gag Glu	ctg Leu	gtg Val 545	aac Asn	cag Gln	atc Ile	att Ile	gag Glu 550	cag Gln	ctg Leu	atc Ile	aag Lys	aag Lys 555	gag Glu	aag Lys	gtg Val	:	1681
tac Tyr	ctg Leu 560	gcc Ala	tgg Trp	gtg Val	cct Pro	gcc Ala 565	càc His	aag Lys	ggc Gly	att Ile	ggg Gly 570	ggc Gly	aat Asn	gag Glu	cag Gln	:	1729

														ctg Leu		•	1777
														aac Asn 605			1825
														aag Lys			1873
														atg Met			1921
														acc Thr			1969
														ggc Gly			2017
														gcc Ala 685			2065
														cac His			2113
														tgc Cys			2161
														cag Gln			2209
cag Gln 735	Gly ggg	gtg Val	gtg Val	gag Glu	tcc Ser 740	atg Met	aac Asn	aag Lys	gag Glu	ctg Leu 745	aag Lys	aag Lys	atc Ile	att Ile	ggg Gly 750		2257
														atg Met 765			2305
gtg Val	ttc Phe	atc Ile	cac His 770	aac Asn	ttc Phe	aag Lys	agg Arg	aag Lys 775	Gly ggg	ggc Gly	atc Ile	Gly ggg	ggc Gly 780	tac Tyr	tcc Ser		2353
														acc Thr			2401
gag Glu	ctc Leu 800	cag Gln	aag Lys	cag Gln	atc Ile	acc Thr 805	aag Lys	atc Ile	cag Gln	aac Asn	ttc Phe 810	agg Arg	gtg Val	tac Tyr	tac Tyr		2449

```
agg gac tcc agg aac ccc ctg tgg aag ggc cct gcc aag ctg ctg tgg
                                                                             2497
Arg Asp Ser Arg Asn Pro Leu Trp Lys Gly Pro Ala Lys Leu Leu Trp
                       820
                                              825
aag ggg gag ggg gct gtg gtg atc cag gac aac tct gac atc aag gtg
Lys Gly Glu Gly Ala Val Val Ile Gln Asp Asn Ser Asp Ile Lys Val
                                                                             2545
                                         840
gtg ccc agg agg aag gcc aag atc atc agg gac tat ggc aag cag atg
                                                                             2593
Val Pro Arg Arg Lys Ala Lys Ile Ile Arg Asp Tyr Gly Lys Gln Met
              850
                                    855
gct ggg gat gac tgt gtg gcc tcc agg cag gat gag gac taa
Ala Gly Asp Asp Cys Val Ala Ser Arg Gln Asp Glu Asp *
                                                                             2635
                                870
                                                                             2650
agcccgggca gatct
<210> 6
<211> 875
<212> PRT
<213> Human Immunodeficiency Virus-1
<400> 6
Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Cys Gly 1 5 10 15
Ala Val Phe Val Ser Pro Ser Glu Ile Ser Ala Pro Ile Ser Pro Ile
20 25 30
Glu Thr Val Pro Val Lys Leu Lys Pro Gly Met Asp Gly Pro Lys Val
                              40
         35
                                                      45
Lys Gln Trp Pro Leu Thr Glu Glu Lys Ile Lys Ala Leu Val Glu Ile
   50
                          55
                                                 60
Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser Lys Ile Gly Pro Glu 65 70 75 80
Asn Pro Tyr Asn Thr Pro Val Phe Ala Ile Lys Lys Lys Asp Ser Thr
                 85
                                        90
Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu Asn Lys Arg Thr Gln 100 105 110
Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His Pro Ala Gly Leu Lys
115 120 125
Lys Lys Ser Val Thr Val Leu Asp Val Gly Asp Ala Tyr Phe Ser
130 140
                                                 140
Val Pro Leu Asp Glu Asp Phe Arg Lys Tyr Thr Ala Phe Thr Ile Pro
                      150
                                             155
                                                                    160
Ser Ile Asn Asn Glu Thr Pro Gly Ile Arg Tyr Gln Tyr Asn Val Leu
165 170 175
                  165
                                        170
Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser Met Thr
180 185 190
Lys Ile Leu Glu Pro Phe Arg Lys Gln Asn Pro Asp Ile Val Ile Tyr
195 200 205
Gln Tyr Met Asp Asp Leu Tyr Val Gly Ser Asp Leu Glu Ile Gly Gln
210 220
                                                  220
His Arg Thr Lys Ile Glu Glu Leu Arg Gln His Leu Leu Arg Trp Gly
                     230
                                           235
                                                                    240
Leu Thr Thr Pro Asp Lys Lys His Gln Lys Glu Pro Pro Phe Leu Trp
245 250 255
Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr Val Gln Pro Ile Val
             260
                                   265
                                                          270
Leu Pro Glu Lys Asp Ser Trp Thr Val Asn Asp Ile Gln Lys Leu Val
275 280 285
Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Pro Gly Ile Lys Val Arg
                           295
```

```
Gln Leu Cys Lys Leu Leu Arg Gly Thr Lys Ala Leu Thr Glu Val Ile
305 310 315 320
                    310
Pro Leu Thr Glu Glu Ala Glu Leu Glu Leu Ala Glu Asn Arg Glu Ile
                325
                                      330
                                                             335
Leu Lys Glu Pro Val His Gly Val Tyr Tyr Asp Pro Ser Lys Asp Leu
340 345 350
Ile Ala Glu Ile Gln Lys Gln Gly Gln Gly Gln Trp Thr Tyr Gln Ile
355 360 365
Tyr Gln Glu Pro Phe Lys Asn Leu Lys Thr Gly Lys Tyr Ala Arg Met
   370
                         375
                                                380
Arg Gly Ala His Thr Asn Asp Val Lys Gln Leu Thr Glu Ala Val Gln 385 390 400
Lys Ile Thr Thr Glu Ser Ile Val Ile Trp Gly Lys Thr Pro Lys Phe
                                      410
                 405
Lys Leu Pro Ile Gln Lys Glu Thr Trp Glu Thr Trp Trp Thr Glu Tyr
                                  425
                                                         430
            420
Trp Gln Ala Thr Trp Ile Pro Glu Trp Glu Phe Val Asn Thr Pro Pro
435
445
                            440
Leu Val Lys Leu Trp Tyr Gln Leu Glu Lys Glu Pro Ile Val Gly Ala
450 460
Glu Thr Phe Tyr Val Asp Gly Ala Ala Asn Arg Glu Thr Lys Leu Gly
465 470 475 480
Lys Ala Gly Tyr Val Thr Asn Arg Gly Arg Gln Lys Val Val Thr Leu
                                      490
                 485
Thr Asp Thr Thr Asn Gln Lys Thr Glu Leu Gln Ala Ile Tyr Leu Ala 500 510
Leu Gln Asp Ser Gly Leu Glu Val Asn Ile Val Thr Asp Ser Gln Tyr 515 525
Ala Leu Gly Ile Ile Gln Ala Gln Pro Asp Gln Ser Glu Ser Glu Leu
530 540
   530
Val Asn Gln Ile Ile Glu Gln Leu Ile Lys Lys Glu Lys Val Tyr Leu
545 550 555 560
Ala Trp Val Pro Ala His Lys Gly Ile Gly Gly Asn Glu Gln Val Asp
565 570 575
                565
Lys Leu Val Ser Ala Gly Ile Arg Lys Val Leu Phe Leu Asp Gly Ile
580 585 590
Asp Lys Ala Gln Asp Glu His Glu Lys Tyr His Ser Asn Trp Arg Ala
       595
                             600
                                                    605
Met Ala Ser Asp Phe Asn Leu Pro Pro Val Val Ala Lys Glu Ile Val
                      615
  610
                                              620
Ala Ser Cys Asp Lys Cys Gln Leu Lys Gly Glu Ala Met His Gly Gln 625 630 635 640
Val Asp Cys Ser Pro Gly Ile Trp Gln Leu Asp Cys Thr His Leu Glu 645 650 655
Gly Lys Val Ile Leu Val Ala Val His Val Ala Ser Gly Tyr Ile Glu
Ala Glu Val Ile Pro Ala Glu Thr Gly Gln Glu Thr Ala Tyr Phe Leu
675 680 685
Leu Lys Leu Ala Gly Arg Trp Pro Val Lys Thr Ile His Thr Asp Asn 690 695 700
Gly Ser Asn Phe Thr Gly Ala Thr Val Arg Ala Ala Cys Trp Trp Ala 705 710 715 720
Gly Ile Lys Gln Glu Phe Gly Ile Pro Tyr Asn Pro Gln Ser Gln Gly 725 730 735
Val Val Glu Ser Met Asn Lys Glu Leu Lys Lys Ile Ile Gly Gln Val 740 745 750
Arg Asp Gln Ala Glu His Leu Lys Thr Ala Val Gln Met Ala Val Phe 755 760 765
Ile His Asn Phe Lys Arg Lys Gly Gly Ile Gly Gly Tyr Ser Ala Gly
                         775
                                                780
Glu Arg Ile Val Asp Ile Ile Ala Thr Asp Ile Gln Thr Lys Glu Leu
                      790
                                                                  800
```

Gln	Lys	Gln	Ile	Thr 805	Lys	Ile	Gln	Asn	Phe 810	Arg	Val	Tyr	Tyr	Arg 815	Asp		
Ser	Arg	Asn	Pro 820		Trp	Lys	Gly	Pro 825	_	Lys	Leu	Leu	Trp 830		Gly		
Glu	Gly	Ala 835		Vaļ	Ile	Gln	Asp 840		Ser	Asp	Ile	Lys 845		Val	Pro		
Arg	Arg 850	Lys	Ala	Lys	Ile	Ile 855		Asp	Tyr	Gly	Lys 860		Met	Ala	Gly		
Asp 865		Cys	Val	Ala	Ser 870		Gln	Asp	Glu	Asp 875	000						
<21:	0> 7 1> 2 2> D 3> H		Imm	.nod	efic:	iency	y Vi:	rus-:	1								
	1> C	DS 8)	. (26:	35)							-						
<400	0> 7																
gato	cacc	atg Met 1			atg Met											. 4	19
tgt Cys 15	gga Gly	gca Ala	gtc Val	ttc Phe	gtt Val 20	tcg Ser	ccc Pro	agc Ser	gag Glu	atc Ile 25	tcc Ser	gcc Ala	ccc Pro	atc Ile	tcc Ser 30	9	97
ccc Pro	att Ile	gag Glu	act Thr	gtg Val 35	cct Pro	gtg Val	Lys	ctg Leu	aag Lys 40	Pro	ggc Gly	atg Met	gat Asp	ggc Gly 45	ccc Pro	14	15
		aag Lys														19	3
gaa Glu	atc Ile	tgc Cys 65	act Thr	gag Glu	atg Met	gag Glu	aag Lys 70	gag Glu	ggc Gly	aaa Lys	atc Ile	tcc Ser 75	aag Lys	att Ile	ggc Gly	24	11
ccc Pro	gag Glu 80	aac Asn	ccc Pro	tac Tyr	aac Asn	acc Thr 85	cct Pro	gtg Val	ttt Phe	gcc Ala	atc Ile 90	aag Lys	aag Lys	aag Lys	gac Asp	28	39
tcc Ser 95	acc Thr	aag Lys	tgg Trp	agg Arg	aag Lys 100	ctg Leu	gtg Val	gac Asp	ttc Phe	agg Arg 105	gag Glu	ctg Leu	aac Asn	aag Lys	agg Arg 110	33	37
acc Thr	cag Gln	gac Asp	ttc Phe	tgg Trp 115	gag Glu	gtg Val	cag Gln	ctg Leu	ggc Gly 120	atc Ile	ccc Pro	cac His	ccc Pro	gct Ala 125	ggc Gly	38	15
ctg Leu	aag Lys	aag Lys	aag Lys 130	aag Lys	tct Ser	gtg Val	act Thr	gtg Val 135	ctg Leu	gct Ala	gtg Val	ggg Gly	gat Asp 140	gcc Ala	tac Tyr	, 43	3
		gtg Val 145														48	1

						gag Glu 165										529
						aag Lys										577
						ccc Pro										625
						gcc Ala										673
						att Ile										721
						gac Asp 245										769
						ctg Leu					Trp				ccc Pro 270	817
						gac Asp										· 865
						tgg Trp										913
						ctg Leu										961
						gag Glu 325										1009
gag Glu 335	atc Ile	ctg Leu	aag Lys	gag Glu	cct Pro 340	gtg Val	cat His	ggg Gly	gtg Val	tac Tyr 345	tat Tyr	gac Asp	ccc Pro	tcc Ser	aag Lys 350	1057
gac Asp	ctg Leu	att Ile	gct Ala	gag Glu 355	atc Ile	cag Gln	aag Lys	cag Gln	ggc Gly 360	cag Gln	ggc Gly	cag Gln	tgg Trp	acc Thr 365	tac Tyr	1105
caa Gln	atc Ile	tac Tyr	cag Gln 370	gag Glu	ccc Pro	ttc Phe	aag Lys	aac Asn 375	ctg Leu	aag Lys	act Thr	gg¢ Gly	aag Lys 380	tat Tyr	gcc Ala	1153
						acc Thr										1201

gtg Val	cag Gln 400	aag Lys	atc Ile	acc Thr	act Thr	gag Glu 405	tcc Ser	att Ile	gtg Val	atc Ile	tgg Trp 410	ggc Gly	aag Lys	acc Thr	ccc Pro		1249
	Phe				atc Ile 420												1297
					acc Thr												1345
					ctg Leu												1393
					tat Tyr												1441
ctg Leu	ggc Gly 480	aag Lys	gct Ala	ggc	tat Tyr	gtg Val 485	acc Thr	aac Asn	agg Arg	ggc Gly	agg Arg 490	cag Gln	aag Lys	gtg Val	gtg Val		1489
					acc Thr 500											•	1537
					tct Ser											*	1585
cag Gln	tat Tyr	gcc Ala	ctg Leu 530	Gly	atc Ile	atč Ile	cag Gln	gcc Ala 535	cag Gln	cct Pro	gat Asp	cag Gln	tct Ser 540	gag Glu	tct Ser	•	1633
gag Glu	ctg Leu	gtg Val 545	aac Asn	cag Gln	atc Ile	att Ile	gag Glu 550	cag Gln	ctg Leu	atc Ile	aag Lys	aag Lys 555	gag Glu	aag Lys	gtg Val	:	1681
tac Tyr	ctg Leu 560	gcc Ala	tgg Trp	gtg Val	cct Pro	gcc Ala 565	cac His	aag Lys	ggc Gly	att Ile	ggg Gly 570	ggc Gly	aat Asn	gag Glu	cag Gln	;	1729
gtg Val 575	Asp	aag Lys	ctg Leu	gtg Val	tct Ser 580	gct Ala	ggc Gly	atc Ile	agg Arg	aag Lys 585	gtg Val	ctg Leu	ttc Phe	ctg Leu	gat Asp 590	•	1777
					cag Gln											:	1825
agg Arg	gct Ala	atg Met	gcc Ala 610	tct Ser	gac Asp	ttc Phe	aac Asn	ctg Leu 615	ccc Pro	cct Pro	gtg Val	gtg Val	gct Ala 620	aag Lys	gag Glu	:	1873
att Ile	gtg Val	gcc Ala 625	tcc Ser	tgt Cys	gac Asp	aag Lys	tgc Cys 630	cag Gln	ctg Leu	aag Lys	ggg Gly	gag Glu 635	gcc Ala	atg Met	cat His	:	1921

ggg Gly	cag Gln 640	gtg Val	gac Asp	tgc Cys	tcc Ser	cct Pro 645	ggc Gly	atc Ile	tgg Trp	cag Gln	ctg Leu 650	gcc Ala	tgc Cys	acc Thr	cac His		1969
ctg Leu 655	gag Glu	ggc Gly	aag Lys	gtg Val	atc Ile 660	ctg Leu	gtg Val	gct Ala	gtg Val	cat His 665	gtg Val	gcc Ala	tcc Ser	ggc Gly	tac Tyr 670		2017
att Ile	gag Glu	gct Ala	gag Glu	gtg Val 675	atc Ile	cct Pro	gct Ala	gag Glu	aca Thr 680	ggc Gly	cag Gln	gag Glu	act Thr	gcc Ala 685	tac Tyr		2065
ttc Phe	ctg Leu	ctg Leu	aag Lys 690	ctg Leu	gct Ala	Gly ggc	agg Arg	tgg Trp 695	cct Pro	gtg Val	aag Lys	acc Thr	atc Ile 700	cac His	act Thr		2113
gcc Ala	aat Asn	ggc Gly 705	tcc Ser	aac Asn	ttc Phe	act Thr	ggg Gly 710	gcc Ala	aca Thr	gtg Val	agg Arg	gct Ala 715	gcc Ala	tgc Cys	tgg Trp		2161
tgg Trp	gct Ala 720	ggc Gly	atc Ile	aag Lys	cag Gln	gag Glu 725	ttt Phe	ggc Gly	atc Ile	ccc Pro	tac Tyr 730	Asn	ccc Pro	cag Gln	tcc Ser		2209
cag Gln 735	ggg ggg	gtg Val	gtg Val	gcc Ala	tcc Ser 740	atg Met	aac Asn	aag Lys	gag Glu	ctg Leu 745	aag Lys	aag Lys	atc Ile	att Ile	ggg Gly 750		2257
cag Gln	gtg Val	agg Arg	gac Asp	cag Gln 755	gct Ala	gag Glu	cac His	ctg Leu	aag Lys 760	aca Thr	gct Ala	gtg Val	cag Gln	atg Met 765	gct Ala	,	2305
gtg Val	ttc Phe	atc Ile	cac His 770	aac Asn	ttc Phe	aag Lys	agg Arg	aag Lys 775	Gly ggg	ggc Gly	atc Ile	Gly	ggc Gly 780	tac Tyr	tcc Ser		2353
gct Ala	ggg Gly	gag Glu 785	agg Arg	att Ile	gtg Val	gac Asp	atc Ile 790	att Ile	gcc Ala	aca Thr	gac Asp	atc Ile 795	cag Gln	acc Thr	aag Lys		2401
gag Glu	ctc Leu 800	cag Gln	aag Lys	cag Gln	atc Ile	acc Thr 805	aag Lys	atc Ile	cag Gln	aac Asn	ttc Phe 810	agg Arg	gtg Val	tac Tyr	tac Tyr		2449
agg Arg 815	gac Asp	tcc Ser	agg Arg	aac Asn	ccc Pro 820	ctg Leu	tgg Trp	aag Lys	ggc Gly	cct Pro 825	gcc Ala	aag Lys	ctg Leu	ctg Leu	tgg Trp 830		2497
aag Lys	ggg Gly	gag Glu	ggg ggg	gct Ala 835	gtg Val	gtg Val	atc Ile	cag Gln	gac Asp 840	aac Asn	tct Ser	gac Asp	atc Ile	aag Lys 845	gtg Val		2545
					gcc Ala												2593
gct Ala	Gly ggg	gat Asp 865	gac Asp	tgt Cys	gtg Val	gcc Ala	tcc Ser 870	agg Arg	cag Gln	gat Asp	gag Glu	gac Asp 875	taa *				2635
agc	ccgg	gca (	gatc	t						•							2650

<210> 8

```
<211> 875
<212> PRT
<213> Human Immunodeficiency Virus-1
Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Cys Gly
                                       10
Ala Val Phe Val Ser Pro Ser Glu Ile Ser Ala Pro Ile Ser Pro Ile
           20
                                 25
Glu Thr Val Pro Val Lys Leu Lys Pro Gly Met Asp Gly Pro Lys Val
                             40
Lys Gln Trp Pro Leu Thr Glu Glu Lys Ile Lys Ala Leu Val Glu Ile
    50
                          55
Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser Lys Ile Gly Pro Glu 65 70 80
Asn Pro Tyr Asn Thr Pro Val Phe Ala Ile Lys Lys Lys Asp Ser Thr 85 90 95
Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu Asn Lys Arg Thr Gln 100 105 110
Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His Pro Ala Gly Leu Lys
                                                    125
     . 115
                              120
Lys Lys Ser Val Thr Val Leu Ala Val Gly Asp Ala Tyr Phe Ser
130 140
Val Pro Leu Asp Glu Asp Phe Arg Lys Tyr Thr Ala Phe Thr Ile Pro 145 150 155 160
Ser Ile Asn Asn Glu Thr Pro Gly Ile Arg Tyr Gln Tyr Asn Val Leu
165 170 175
                                                             175
Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser Met Thr
180 185 190
            180
                                  185
                                                        190
Lys Ile Leu Glu Pro Phe Arg Lys Gln Asn Pro Asp Ile Val Ile Tyr 195 .200 205
Gln Tyr Met Ala Ala Leu Tyr Val Gly Ser Asp Leu Glu Ile Gly Gln 210 215 220
His Arg Thr Lys Ile Glu Glu Leu Arg Gln His Leu Leu Arg Trp Gly
                     230
225
                                           235
Leu Thr Thr Pro Asp Lys Lys His Gln Lys Glu Pro Pro Phe Leu Trp 245 250 255
Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr Val Gln Pro Ile Val 260 265 270
Leu Pro Glu Lys Asp Ser Trp Thr Val Asn Asp Ile Gln Lys Leu Val 275 280 285
Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Pro Gly Ile Lys Val Arg
                         295
Gln Leu Cys Lys Leu Leu Arg Gly Thr Lys Ala Leu Thr Glu Val Ile 305 310 315 320
Pro Leu Thr Glu Glu Ala Glu Leu Glu Leu Ala Glu Asn Arg Glu Ile
                 325
                                       330
Leu Lys Glu Pro Val His Gly Val Tyr Tyr Asp Pro Ser Lys Asp Leu 340 345
Ile Ala Glu Ile Gln Lys Gln Gly Gln Gly Gln Trp Thr Tyr Gln Ile 355 360 365
Tyr Gln Glu Pro Phe Lys Asn Leu Lys Thr Gly Lys Tyr Ala Arg Met 370 380
Arg Gly Ala His Thr Asn Asp Val Lys Gln Leu Thr Glu Ala Val Gln 385 390 395 400
Lys Ile Thr Thr Glu Ser Ile Val Ile Trp Gly Lys Thr Pro Lys Phe
                 405
                                      410
Lys Leu Pro Ile Gln Lys Glu Thr Trp Glu Thr Trp Trp Thr Glu Tyr 420 425 430
Trp Gln Ala Thr Trp Ile Pro Glu Trp Glu Phe Val Asn Thr Pro Pro
```

PCT/US00/34724 WO 01/45748

```
Leu Val Lys Leu Trp Tyr Gln Leu Glu Lys Glu Pro Ile Val Gly Ala
    450
                        455
                                              460
Glu Thr Phe Tyr Val Ala Gly Ala Ala Asn Arg Glu Thr Lys Leu Gly
                    470
                                         475
Lys Ala Gly Tyr Val Thr Asn Arg Gly Arg Gln Lys Val Val Thr Leu
                485
                                      490
Thr Asp Thr Thr Asn Gln Lys Thr Ala Leu Gln Ala Ile Tyr Leu Ala 500 510
                                 505
Leu Gln Asp Ser Gly Leu Glu Val Asn Ile Val Thr Ala Ser Gln Tyr
                            520
Ala Leu Gly Ile Ile Gln Ala Gln Pro Asp Gln Ser Glu Ser Glu Leu
                        535
                                              540
Val Asn Gln Ile Ile Glu Gln Leu Ile Lys Lys Glu Lys Val Tyr Leu 545 550 555 560
                                          555
545
                    550
Ala Trp Val Pro Ala His Lys Gly Ile Gly Gly Asn Glu Gln Val Asp
565 570
                565
Lys Leu Val Ser Ala Gly Ile Arg Lys Val Leu Phe Leu Asp Gly Ile
580 585 590
Asp Lys Ala Gln Asp Glu His Glu Lys Tyr His Ser Asn Trp Arg Ala
595 600 605
        595
                             600
                                                  605
Met Ala Ser Asp Phe Asn Leu Pro Pro Val Val Ala Lys Glu Ile Val
   610
                        615
                                              620
Ala Ser Cys Asp Lys Cys Gln Leu Lys Gly Glu Ala Met His Gly Gln 625 630 635 640
                    630
                                         635
Val Asp Cys Ser Pro Gly Ile Trp Gln Leu Ala Cys Thr His Leu Glu 645 650 655
Gly Lys Val Ile Leu Val Ala Val His Val Ala Ser Gly Tyr Ile Glu 660 665 670
                                 665
Ala Glu Val Ile Pro Ala Glu Thr Gly Gln Glu Thr Ala Tyr Phe Leu
675 680 685
Leu Lys Leu Ala Gly Arg Trp Pro Val Lys Thr Ile His Thr Ala Asn 690 695 700
Gly Ser Asn Phe Thr Gly Ala Thr Val Arg Ala Ala Cys Trp Trp Ala 705 710 715 720
Gly Ile Lys Gln Glu Phe Gly Ile Pro Tyr Asn Pro Gln Ser Gln Gly 725 730 735
Val Val Ala Ser Met Asn Lys Glu Leu Lys Lys Ile Ile Gly Gln Val
            740
                            745
                                                      750
Arg Asp Gln Ala Glu His Leu Lys Thr Ala Val Gln Met Ala Val Phe 755 760 765
Ile His Asn Phe Lys Arg Lys Gly Gly Ile Gly Gly Tyr Ser Ala Gly 770 780
Glu Arg Ile Val Asp Ile Ile Ala Thr Asp Ile Gln Thr Lys Glu Leu
                     790
785
                                          795
Gln Lys Gln Ile Thr Lys Ile Gln Asn Phe Arg Val Tyr Tyr Arg Asp
805 815
                                     810
                805
Ser Arg Asn Pro Leu Trp Lys Gly Pro Ala Lys Leu Leu Trp Lys Gly 820 825 830
                                                    830
Glu Gly Ala Val Val Ile Gln Asp Asn Ser Asp Ile Lys Val Val Pro
                            840
       835
                                                 845
Arg Arg Lys Ala Lys Ile Ile Arg Asp Tyr Gly Lys Gln Met Ala Gly
                       855
   850
                                             860
Asp Asp Cys Val Ala Ser Arg Gln Asp Glu Asp
865
                     870
<210> 9
<211> 4945
```

tegegegttt eggtgatgae ggtgaaaaee tetgacaeat geageteeeg gagaeggtea

<sup>&</sup>lt;212> DNA <213> E. coli (V1Jns-tpa)

<sup>&</sup>lt;400> 9

cagcttgtct	gtaagcggat	gccgggagca	gacaagcccg	tcagggcgcg	tcagcgggtg	120
ttggcgggtg	teggggetgg	cttaactato	cggcatcaga	gcagattgta	ctgagagtgc	180
ccggcgggcg	ccggggccgg	cccaaccacg		900900000	25000550	240
accatatgcg	gtgtgaaata	ccgcacagat	gcgtaaggag	aaaataccyc	accayacty	
ctattggcca	ttgcatacgt	totatccata	tcataatatg	tacatttata	ttggctcatg	300
tagaagatta	ccccatatt	gagattgatt	attgactagt	tattaatagt	aatcaattac	360
LCCAACACCA	ccgccacgcc	gacartgatt	accoaccage	cactaatage		
ggggtcatta	gttcatagcc	catatatgga	gttccgcgtt	acataactta	cggtaaatgg	420
cccacctage	tgaccgccca	acqacccccq	cccattgacg	tcaataatga	catatattcc	480
		atttaatta	acgtcaatgg	atagagtatt	taccotaaac	540
catagtaacg	ccaatayyya	Cittecatty	acgicaatgg	gradagrare	cacygraaac	
tgcccacttg	gcagtacatc	aagtgtatca	tatgccaagt	acgcccccta	ttgacgtcaa	600
tgacggtaaa	taggegget	ggcattatgc	ccagtacatg	accttatggg	actttcctac	660
		5500000000	tattaggatg	atastacaat	tttaagaagta	720
Liggiagiac	acceacycac	tagtcatege	tattaccatg	grgargrggr	cccggcagca	
catcaatggg	cgtggatagc	ggtttgactc	acggggattt	ccaagtctcc	accccattga	780
catcaataga	agtttgttt	ggcaccaaaa	tcaacgggac	tttccaaaat	gtcgtaacaa	840
chacacaca	ttanaaan	tagacaataa	gcgtgtacgg	taggaggtet	atataancan	900
CLCCGCCCCA	Ligacycada	cgggcggcag	gcgcgcacgg	cgggaggccc	acacaageag	
agctcgttta	gtgaaccgtc	agatcgcctg	gagacgccat	ccacgctgtt	ttgacctcca	960
tagaagacac	coogaccoat	ccagcctccg	cggccgggaa	cggtgcattg	gaacgcggat	1020
toggagtace	asasatasca	taagtaccgc	ctatagactc	tatagggaga	cccctttagc	1080
			gggcctatac			1140
taggtgatgg	tatagettag	cctataggtg	tgggttattg	accattattg	accactcccc	1200
tattaataac	gatactttcc	attactaatc	cataacatgg	ctctttacca	caactatete	1260
tattggtgat	gatactetee	h-b-b-abba			****	
tattggctat	atgccaatac	tetgteette	agagactgac	acygactetg	Lattituda	1320
ggatggggtc	ccatttatta	tttacaaatt	cacatataca	acaacgccgt	cccccgtgcc	1380
			tccacgcgaa			1440
cycaycccc	actadacuta	gegegggace	tecacgegaa		gegeeeegga	1500
catgggctct	tctccggtag	cggcggagct	tccacatccg	agecetggte	ccatgcctcc	
agcggctcat	gategetegg	cagctccttg	ctcctaacag	tggaggccag	acttaggcac	1560
			cacaaggccg			. 1620
agcacaacgc			~~*~~~~	0550550055	taagggggg	1680
gaaaatgage	grggagarrg	ggetegeaeg	gctgacgcag	atygaagact	caayycaycy	
gcagaagaag	atgcaggcag	ctgagttgtt	gtattctgat	aagagtcaga	ggtaactccc	1740
attacaatac	tattaacaat	ggaggggagt	gtagtctgag	cagtactcgt	tactaccaca	1800
geogeggege		2242222424	Souscostat	tagtttagst	acatatttta	
					gggtcttttc	1000
			gcaatgaaga			1920
ctactatata	gagcagtctt	catttcaccc	agcgagatct	actatacett	ctagttgcca	1980
gagatatatt	atttaggggt	caccataca	ttccttgacc	ctagaaggta	ccactcccac	2040
gecatetytt	geeegeeeee	cccccgcgcc	LLCCLLgacc	ctggaaggtg		
tgtcctttcc	taataaaatg	aggaaattgc	atcgcattgt	ctgagtaggt	gtcattctat	2100
tetaaaaaat	aaaataaaac	aggacagcaa	gggggaggat	tgggaagaca	atagcaggca	2160
			tgcggccagg			2220
cgccggggac	geggegggee	ccacggccgc	cgcggccagg	cgccgaagaa	-t-t	
tectectggg	ccagaaagaa	gcaggcacat	ccccttctct	gtgacacacc	ctgtccacgc	2280
ccctaattct	tagttccagc	cccactcata	ggacactcat	agctcaggag	ggctccgcct	2340
			gtctctccct			2400
caaacctagc	ctccaagagt	gggaagaat	taaagcaaga	taggetatta	agtgcagagg	2460
gagagaaaat	gcctccaaca	tgtgaggaag	taatgagaga	aatcatagaa	tttcttccgc	2520
			cgttcggctg			2580
						2640
			atcaggggat			
			taaaaaggcc			2700
taggeteege	cccctgacg	agcatcacaa	aaatcgacgc	tcaaqtcaqa	ggtggcgaaa	2760
22222222	ctataaacat	accaddocatt	tcccctgga	accteceted	tacactatac	2820
tgttccgacc	ctgccgctta	ccggatacct	gtccgccttt	crecerregg	gaagegtgge	2880
			cagttcggtg			2940
			cgaccgctgc			3000
tettgagtee	aacccggtaa	gacacgactt	atcgccactg	gcagcagcca	ctggtaacag	3060
gattagcaga	gcgaggtatg	taggcggtgc	tacagagttc	ttgaagtggt	ggcctaacta	3120
coactacact	24224224	tatttggtat	ctgcgctctg	ctgaagccag	ttaccttcoo	3180
			acaaaccacc			3240
tgtttqcaaq	cagcagatta	cgcgcagaaa	aaaaggatct	caagaagatc	ctttgatctt	3300
			aaactcacgt			3360
			tttaaattaa			3420
ctaaagtata	tatgagtaaa	cttggtctga	cagttaccaa	tgcttaatca	gtgaggcacc	3480
			catagttgcc			3540
						3600
			tgactcatac			
atccagccag	aaagtgaggg	agccacggtt	gatgagagct	ttgttgtagg	tggaccagtt	3660
ggtgattttg	aacttttgct	ttgccacaga	acggtctgcg	ttatcaaaaa	gatgcgtgat	3720
			tattcaacaa			3780
Gugaluctuc	aacccaycad	aaytttyatt	carccaacaa	agecyclytt	cogreaged	3700

```
agcgtaatgc tctgccagtg ttacaaccaa ttaaccaatt ctgattagaa aaactcatcg
agcatcaaat gaaactgcaa tttattcata tcaggattat caataccata tttttgaaaa agccgtttct gtaatgaagg agaaaactca ccgaggcagt tccataggat ggcaagatcc
                                                                                       3900
                                                                                       3960
tggtatcggt ctgcgattcc gactcgtcca acatcaatac aacctattaa tttcccctcg
                                                                                       4020
tcaaaaataa ggttatcaag tgagaaatca ccatgagtga cgactgaatc cggtgagaat ggcaaaagct tatgcatttc tttccagact tgttcaacag gccagccatt acgctcgtca
                                                                                       4080
                                                                                       4140
tcaaaatcac tcgcatcaac caaaccgtta ttcattcgtg attgcgcctg agcgagacga
                                                                                       4200
aatacgcgat cgctgttaaa aggacaatta caaacaggaa tcgaatgcaa ccggcgcagg
                                                                                       4260
aacactgcca gcgcatcaac aatattttca cctgaatcag gatattcttc taatacctgg
                                                                                       4320
aatgctgttt tcccggggat cgcagtggtg agtaaccatg catcatcagg agtacggata aaatgcttga tggtcggaag aggcataaat tccgtcagcc agtttagtct gaccatctca tctgtaacat cattggcaac gctacctttg ccatgttca gaaacaactc tggcgcatcg
                                                                                       4380
                                                                                       4440
                                                                                       4500
ggcttcccat acaatcgata gattgtcgca cctgattgcc cgacattatc gcgagcccat
                                                                                       4560
ttatacccat ataaatcagc atccatgttg gaatttaatc gcggcctcga gcaagacgtt tcccgttgaa tatggctcat aacaccctt gtattactgt ttatgtaagc agacagtttt
                                                                                       4620
                                                                                       4680
attgttcatg atgatatatt tttatcttgt gcaatgtaac atcagagatt ttgagacaca
                                                                                       4740
acgiggetti ecceccece ceattattga ageatttate agggttattg teteatgage
                                                                                       4800
ggatacatat ttgaatgtat ttagaaaaat aaacaaatag gggttccgcg cacatttccc
                                                                                       4860
cgaaaagtgc cacctgacgt ctaagaaacc attattatca tgacattaac ctataaaaaat
                                                                                       4920
aggcgtatca cgaggccctt tcgtc
                                                                                       4945
<210> 10
<211> 23
<212> DNA
<213> Artificial Sequence
<220>
<223> oligonucleotide
<400> 10
                                                                                         23
ctatataagc agagctcgtt tag
<210> 11
<211> 30
<212> DNA
<213> Artificial Sequence
<220>
<223> oligonucleotide
<400> 11
gtagcaaaga tctaaggacg gtgactgcag
                                                                                         30
<210> 12
<211> 39
<212> DNA
<213> Artificial Sequence
<220>
<223> oligonucleotide
<400> 12
gtatgtgtct gaaaatgagc gtggagattg ggctcgcac
                                                                                         39
<210> 13
<211> 39
<212> DNA
<213> Artificial Sequence
<220>
<223> oligonucleotide
```

```
<400> 13
                                                                                                 39
gtgcgagccc aatctccacg ctcattttca gacacatac
<210> 14
<211> 4432
<212> DNA
<213> E. coli (V1J plasmid)
<400> 14
tcgcgcgttt cggtgatgac ggtgaaaacc tctgacacat gcagctcccg gagacggtca
                                                                                                 60
cagettgtet gtaageggat geegggagea gacaageeeg teagggegeg teagegggtg
                                                                                                120
ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc
                                                                                                180
accatatgcg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcagattgg
                                                                                                240
ctattggcca ttgcatacgt tgtatccata tcataatatg tacatttata ttggctcatg tccaacatta ccgccatgtt gacattgatt attgactagt tattaatagt aatcaattac
                                                                                                300
                                                                                                360
ggggtcatta gttcatagcc catatatgga gttccgcgtt acataactta cggtaaatgg
                                                                                                420
cccgcctggc tgaccgccca acgacccccg cccattgacg tcaataatga cgtatgttcc
                                                                                                480
catagtaacg ccaataggga ctttccattg acgtcaatgg gtggagtatt tacggtaaac
                                                                                               540
tgcccacttg gcagtacatc aagtgtatca tatgccaagt acgcccccta ttgacgtcaa
                                                                                                600
tgacggtaaa tggcccgcct ggcattatgc ccagtacatg accttatggg actttcctac ttggcagtac atctacgtat tagtcatcgc tattaccatg gtgatgcggt tttggcagta
                                                                                                660
                                                                                                720
catcaatggg cgtggatagc ggtttgactc acggggattt ccaagtctcc accccattga
                                                                                                780
cgtcaatggg agtttgtttt ggcaccaaaa tcaacgggac tttccaaaat gtcgtaacaa
                                                                                               840
ctccgccca ttgacgcaaa tgggcggtag gcgtgtacgg tgggaggtct atataagcag agctcgttta gtgaaccgtc agatcgcctg gagacgccat ccacgctgtt ttgacctca tagaagacac cgggaccgat ccagcctccg cggccgggaa cggtgcattg gaacgcggt
                                                                                                900
                                                                                               960
                                                                                              1020
teccegtgee aagagtgaeg taagtaeege etatagagte tataggeeea ecceettgge
                                                                                              1080
ttcttatgca tgctatactg tttttggctt ggggtctata cacccccgct tcctcatgtt
                                                                                              1140
ataggtgatg gtatagctta gcctataggt gtgggttatt gaccattatt gaccactccc
                                                                                              1200
ctattggtga cgatactttc cattactaat ccataacatg gctctttgcc acaactctct
                                                                                              1260
ttattggcta tatgccaata cactgtcctt cagagactga cacggactct gtatttttac
                                                                                              1320
aggatgggt ctcatttatt atttacaaat tcacatatac aacaccaccg tccccagtgc
                                                                                              1380
ccgcagtttt tattaaacat aacgtgggat ctccacgcga atctcgggta cgtgttccgg
                                                                                              1440
acatgggctc ttctccggta gcggcggagc ttctacatcc gagccctgct cccatgcctc
                                                                                              1500
cagogactca tggtcgctcg gcagctcctt gctcctaaca gtggaggcca gacttaggca
                                                                                              1560
cagcacgatg cccaccacca ccagtgtgcc gcacaaggcc gtggcggtag ggtatgtgtc tgaaaatgag ctcggggagc gggcttgcac cgctgacgca tttggaagac ttaaggcagc
                                                                                              1620
                                                                                              1680
ggcagaagaa gatgcaggca gctgagttgt tgtgttctga taagagtcag aggtaactcc
                                                                                              1740
cgttgcggtg ctgttaacgg tggagggcag tgtagtctga gcagtactcg ttgctgccgc gcgcgccacc agacataata gctgacagac taacagactg ttcctttcca tgggtcttt
                                                                                              1800
                                                                                              1860
ctgcagtcac cgtccttaga tctgctgtgc cttctagttg ccagccatct gttgtttgcc
                                                                                              1920
cctccccgt gccttccttg accctggaag gtgccactcc cactgtcctt tcctaataaa atgaggaaat tgcatcgcat tgtctgagta ggtgtcattc tattctgggg ggtgggtgg
                                                                                              1980
                                                                                              2040
ggcagcacag caagggggag gattgggaag acaatagcag gcatgctggg gatgcggtgg gctctatggg tacccaggtg ctgaagaatt gacccggttc ctcctgggcc agaaagaagc aggcacatcc ccttctctgt gacacaccct gtccacgccc ctggttctta gttccagccc
                                                                                              2100
                                                                                              2160
                                                                                              2220
cactcatagg acactcatag ctcaggaggg ctccgccttc aatcccaccc gctaaagtac ttggaggggt ctctcctcc ctcatcagcc caccaaacca aacctagcct ccaagagtgg
                                                                                              2280
                                                                                              2340
gaagaaatta aagcaagata ggctattaag tgcagaggga gagaaaatgc ctccaacatg
                                                                                              2400
tgaggaagta atgagagaaa tcatagaatt tcttccgctt cctcgctcac tgactcgctg cgctcggtcg ttcggctgcg gcgagcggta tcagctcact caaaggcggt aatacggtta
                                                                                              2460
                                                                                              2520
tccacagaat caggggataa cgcaggaaag aacatgtgag caaaaggcca gcaaaaggcc
                                                                                              2580
aggaaccgta aaaaggccgc gttgctggcg tttttccata ggctccgccc ccctgacgag catcacaaaa atcgacgctc aagtcagagg tggcgaaacc cgacaggact ataaagatac
                                                                                              2640
                                                                                              2700
caggogtttc cccctggaag ctccctcgtg cgctctcctg ttccgaccct gccgcttacc
                                                                                              2760
ggatacetgt cegeetttet ecetteggga agegtggege ttteteaatg eteacgetgt aggtatetea gtteggtgta ggtegttege tecaagetgg getgtgtgea egaaceeee
                                                                                              2820
                                                                                              2880
gttcagcccg accgctgcgc cttatccggt aactatcgtc ttgagtccaa cccggtaaga
                                                                                              2940
cacgacttat cgccactggc agcagccact ggtaacagga ttagcagagc gaggtatgta ggcggtgcta cagagttctt gaagtggtgg cctaactacg gctacactag aaggacagta
                                                                                              3000
                                                                                              3060
tttggtatct gegetetget gaagecagtt acetteggaa aaagagttgg tagetettga
teeggeaaac aaaceacege tggtageggt ggtttttttg tttgeaagea geagattaeg
                                                                                              3120
                                                                                              3180
```

```
3240
cgcagaaaaa aaggatctca agaagatcct ttgatctttt ctacggggtc tgacgctcag
tggaacgaaa actcacgtta agggattttg gtcatgagat tatcaaaaag gatcttcacc tagatccttt taaattaaaa atgaagttt aaatcaatct aaagtatata tgagtaaact
                                                                                    3300
                                                                                    3360
tggtctgaca gttaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtctattt
                                                                                    3420
cgttcatcca tagttgcctg actccccgtc gtgtagataa ctacgatacg ggagggctta ccatctggcc ccagtgctgc aatgataccg cgagacccac gctcaccggc tccagattta
                                                                                    3480
                                                                                    3540
tcagcaataa accagccagc cggaagggcc gagcgcagaa gtggtcctgc aactttatcc
                                                                                    3600
gcctcatcc agtctattaa ttgttgccgg gaagctagag taagtagttc gccagttaat agtttgcgca acgttgttgc cattgctaca ggcatcgtgg tgtcacgctc gtcgtttggt
                                                                                    3660
                                                                                    3720
atggetteat teageteegg tteceaacga teaaggegag ttacatgate ecceatgttg
                                                                                    3780
tgcaaaaaag cggttagctc cttcggtcct ccgatcgttg tcagaagtaa gttggccgca gtgttatcac tcatggttat ggcagcactg cataattctc ttactgtcat gccatccgta
                                                                                    3840
                                                                                    3900
                                                                                    3960
agatgetttt etgtgaetgg tgagtaetea accaagteat tetgagaata gtgtatgegg
cgaccgagtt gctcttgccc ggcgtcaata cgggataata ccgcgccaca tagcagaact
                                                                                    4020
ttaaaagtgc tcatcattgg aaaacgttct tcggggcgaa aactctcaag gatcttaccg
                                                                                    4080
ctgttgagat ccagttcgat gtaacccact cgtgcaccca actgatcttc agcatctttt
                                                                                    4140
acttcacca gcgtttctgg gtgagcaaaa acaggaaggc aaaatgccgc aaaaaaggga ataagggcga cacggaaatg ttgaatactc atactctcc tttttcaata ttattgaagc
                                                                                    4200
                                                                                    4260
attatcagg gttattgtct catgagegga tacatatttg aatgtattta gaaaaataaa caaatagggg tteegegeac attteeeega aaagtgeeac etgaegteta agaaaceatt
                                                                                    4320
                                                                                    4380
attatcatga cattaaccta taaaaatagg cgtatcacga ggccctttcg tc
                                                                                    4432
<210> 15
<211> 4864
<212> DNA
<213> E. coli (VlJneo plasmid)
tegegegttt eggtgatgae ggtgaaaace tetgacacat geageteeeg gagaeggtea
                                                                                      120
cagettgtet gtaageggat geegggagea gacaageeeg teagegggeg teagegggtg
ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc
                                                                                      180
accatatgcg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcagattgg
                                                                                      240
ctattggcca ttgcatacgt tgtatccata tcataatatg tacatttata ttggctcatg tccaacatta ccgccatgtt gacattgatt attgactagt tattaatagt aatcaattac
                                                                                      300
                                                                                      360
ggggtcatta gttcatagcc catatatgga gttccgcgtt acataactta cggtaaatgg
                                                                                      420
cocgcotgge tgaccgccca acgacccccg cccattgacg tcaataatga cgtatgttcc catagtaacg ccaataggga ctttccattg acgtcaatgg gtggagtatt tacggtaaac
                                                                                      480
                                                                                     540
tgcccacttg gcagtacatc aagtgtatca tatgccaagt acgccccta ttgacgtcaa
                                                                                      600
tgacggtaaa tggcccgcct ggcattatgc ccagtacatg accttatggg actttcctac ttggcagtac atctacgtat tagtcatcgc tattaccatg gtgatgcggt tttggcagta
                                                                                      660
                                                                                      720
catcaatggg cgtggatagc ggtttgactc acggggattt ccaagtctcc accccattga
                                                                                      780
cgtcaatggg agtttgtttt ggcaccaaaa tcaacgggac tttccaaaat gtcgtaacaa
                                                                                     840
ctccgcccca ttgacgcaaa tgggcggtag gcgtgtacgg tgggaggtct atataagcag
                                                                                     900
agetegttta gtgaacegte agategeetg gagacgeeat ceaegetgtt ttgaceteea
                                                                                     960
tagaagacac cgggaccgat ccagcctccg cggccgggaa cggtgcattg gaacgcggat
                                                                                    1020
teccegtgee aagagtgacg taagtacege ctatagagte tataggeeca eccettgge
                                                                                    1080
ttcttatgca tgctatactg tttttggctt ggggtctata caccccgct tcctcatgtt
                                                                                    1140
ataggtgatg gtatagctta gcctataggt gtgggttatt gaccattatt gaccactccc
                                                                                    1200
ctattggtga cgatactttc cattactaat ccataacatg gctctttgcc acaactctct
                                                                                    1260
ttattggcta tatgccaata cactgtcctt cagagactga cacggactct gtattttac aggatggggt ctcatttatt atttacaaat tcacatatac aacaccaccg tccccagtgc
                                                                                    1320
                                                                                    1380
ccgcagtttt tattaaacat aacgtgggat ctccacgcga atctcgggta cgtgttccgg
                                                                                    1440
acatgggete tteteeggta geggeggage ttetacatee gagecetget eccatgeete cagegactea tggtegeteg geageteett geteetaaca gtggaggeca gaettaggea
                                                                                    1500
                                                                                    1560
cagcacgatg cccaccacca ccagtgtgcc gcacaaggcc gtggcggtag ggtatgtgtc
                                                                                    1620
tgaaaatgag ctcggggagc gggcttgcac cgctgacgca tttggaagac ttaaggcagc
                                                                                    1680
ggcagaagaa gatgcaggca gctgagttgt tgtgttctga taagagtcag aggtaactcc
                                                                                    1740
cgttgcggtg ctgttaacgg tggagggcag tgtagtctga gcagtactcg ttgctgccgc
                                                                                    1800
gegegecace agacataata getgacagae taacagaetg tteettteca tgggtetttt
                                                                                    1860
ctgcagtcac cgtccttaga tctgctgtgc cttctagttg ccagccatct gttgtttgcc
                                                                                    1920
1980
                                                                                    2040
```

```
ggcagcacag caagggggag gattgggaag acaatagcag gcatgctggg gatgcggtgg
                                                                                      2100
getetatggg tacceaggtg etgaagaatt gaceeggtte eteetgggee agaaagaage
                                                                                      2160
aggcacatcc cettetetgt gacacaccet gtccacgccc etggttetta gttccagccc cactcatagg acactcatag etcaggaggg etcegeette aatcccacce getaaagtac
                                                                                      2220
                                                                                      2280
ttggagcggt ctctccctcc ctcatcagcc caccaaacca aacctagcct ccaagagtgg
                                                                                      2340
gaagaaatta aagcaagata ggctattaag tgcagaggga gagaaaatgc ctccaacatg tgaggaagta atgagagaaa tcatagaatt tcttccgctt cctcgctcac tgactcgctg
                                                                                      2400
                                                                                      2460
                                                                                      2520
cgctcggtcg ttcggctgcg gcgagcggta tcagctcact caaaggcggt aatacggtta
tccacagaat caggggataa cgcaggaaag aacatgtgag caaaaggcca gcaaaaggcc
                                                                                      2580
aggaaccgta aaaaggccgc gttgctggcg tttttccata ggctccgccc ccctgacgag
                                                                                      2640
catcacaaaa atcgacgctc aagtcagagg tggcgaaacc cgacaggact ataaagatac
                                                                                      2700
                                                                                      2760
caggegtttc cccctggaag ctccctcgtg cgctctcctg ttccgaccct gccgcttacc
ggatacctgt ccgcctttct cccttcggga agcgtggcgc tttctcaatg ctcacgctgt
                                                                                      2820
aggtatctca gttcggtgta ggtcgttcgc tccaagctgg gctgtgtgca cgaacccc gttcagcccg accgctgcgc cttatccggt aactatcgtc ttgagtccaa cccggtaaga
                                                                                      2880
                                                                                      2940
cacgacttat cgccactggc agcagccact ggtaacagga ttagcagagc gaggtatgta
                                                                                      3000
ggcggtgcta cagagttctt gaagtggtgg cctaactacg gctacactag aaggacagta tttggtatct gcgctctgct gaagccagtt accttcggaa aaagagttgg tagctcttga
                                                                                      3060
                                                                                      3120
tccggcaaac aaaccaccgc tggtagcggt ggtttttttg tttgcaagca gcagattacg cgcagaaaaa aaggatctca agaagatcct ttgatctttt ctacggggtc tgacgctcag
                                                                                      3180
                                                                                      3240
tggaacgaaa actcacgtta agggattttg gtcatgagat tatcaaaaag gatcttcacc
                                                                                      3300
                                                                                      3360
tagatccttt taaattaaaa atgaagtttt aaatcaatct aaagtatata tgagtaaact
tggtctgaca gttaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtctattt
                                                                                      3420
cgttcatcca tagttgcctg actccggggg gggggggcgc tgaggtctgc ctcgtgaaga aggtgttgct gactcatacc aggcctgaat cgccccatca tccagccaga aagtgaggga
                                                                                      3480
                                                                                      3540
gccacggttg atgagagctt tgttgtaggt ggaccagttg gtgattttga acttttgctt
                                                                                      3600
tgccacggaa cggtctgcgt tgtcgggaag atgcgtgatc tgatccttca actcagcaaa
                                                                                      3660
agttcgattt attcaacaaa gccgccgtcc cgtcaagtca gcgtaatgct ctgccagtgt
                                                                                      3720
tacaaccaat taaccaattc tgattagaaa aactcatcga gcatcaaatg aaactgcaat
                                                                                      3780
ttattcatat caggattatc aataccatat ttttgaaaaa gccgtttctg taatgaagga
                                                                                      3840
gaaaactcac cgaggcagtt ccataggatg gcaagatcct ggtatcggtc tgcgattccg actcgtccaa catcaataca acctattaat ttcccctcgt caaaaataag gttatcaagt
                                                                                      3900
                                                                                      3960
gagaaatcac catgagtgac gactgaatcc ggtgagaatg gcaaaagctt atgcatttct
                                                                                      4020
ttccagactt gttcaacagg ccagccatta cgctcgtcat caaaatcact cgcatcaacc aaaccgttat tcattcgtga ttgcgcctga gcgagacgaa atacgcgatc gctgttaaaa
                                                                                      4080
                                                                                      4140
ggacaattac aaacaggaat cgaatgcaac cggcgcagga acactgccag cgcatcaaca
                                                                                      4200
atattttcac ctgaatcagg atattcttct aatacctgga atgctgtttt cccgggggatc
                                                                                      4260
gcagtggtga gtaaccatgc atcatcagga gtacggataa aatgcttgat ggtcggaaga
                                                                                      4320
ggcataaatt ccgtcagcca gtttagtctg accatctcat ctgtaacatc attggcaacg
                                                                                      4380
ctaccttige catgittcag aaacaactct ggegeategg getteceata caategatag attgtegeae etgattgeee gacattateg egageeeatt tatacceata taaateagea
                                                                                      4440
                                                                                      4500
tccatgttgg aatttaatcg cggcctcgag caagacgttt cccgttgaat atggctcata acacccttg tattactgtt tatgtaagca gacagtttta ttgttcatga tgatatattt
                                                                                      4560
                                                                                      4620
ttatcttgtg caatgtaaca tcagagattt tgagacacaa cgtggctttc ccccccccc
                                                                                      4680
cattattgaa gcatttatca gggttattgt ctcatgagcg gatacatatt tgaatgtatt tagaaaaata aacaaatagg ggttccgcgc acatttcccc gaaaagtgcc acctgacgtc
                                                                                      4740
                                                                                      4800
taagaaacca ttattatcat gacattaacc tataaaaata ggcgtatcac gaggcccttt
                                                                                      4860
                                                                                      4864
<210> 16
<211> 4867
<212> DNA
<213> E. coli (V1Jns plasmid)
tcgcgcgttt cggtgatgac ggtgaaaacc tctgacacat gcagctcccg gagacggtca
                                                                                        60
cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg
                                                                                       120
ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc
                                                                                       180
accatatgcg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcagattgg
                                                                                       240
ctattggcca ttgcatacgt tgtatccata tcataatatg tacatttata ttggctcatg
                                                                                       300
tccaacatta ccgccatgtt gacattgatt attgactagt tattaatagt aatcaattac
                                                                                       360
ggggtcatta gttcatagcc catatatgga gttccgcgtt acataactta cggtaaatgg
                                                                                       420
```

					-	
cccacctaac	tgaccgccca	acgacccccg	cccattgacg	tcaataatga	cgtatgttcc	480
catactaacc	ccaataggga	ctttccattg	acatcaataa	gtggagtatt	tacqqtaaac	540
tatagtaatg	ccaacaggga	cccccatty	acyccaacyy	gragageace	thesetees	
tgcccacttg	gcagtacatc	aagtgtatca	tatgccaagt	acgcccccta	ttgacgtcaa	600
tgacggtaaa	tggcccgcct	ggcattatgc	ccagtacatg	accttatggg	actttcctac	660
ttggcagtac	atctacotat	tagtcatcgc	tattaccato	ataatacaat	tttggcagta	720
catcaatooo	catagataga	ggtttgactc	acqqqqattt	ccaagtetee	accccattga	780
cattaatggg	cytygatayc	ggtttgattt	acggggaccc	CCaagccccc	accecactga	
		ggcaccaaaa				840
ctccgcccca	ttgacgcaaa	tgggcggtag	gcgtgtacgg	tgggaggtct	atataagcag	900
agctcgttta	gtgaaccgtc	agatcgcctg	gagacgccat	ccacactatt	ttgacctcca	960
		ccagcctccg				1020
		taagtaccgc				1080
		ttttggcttg				1140
taggtgatgg	tatagcttag	cctataggtg	tgggttattg	accattattg	accactcccc	1200
		attactaatc				1260
		tctgtccttc				1320
		tttacaaatt				1380
		gcgtgggatc				1440
catgggctct	tctccggtag	cggcggagct	tccacatccg	agccctggtc	ccatgcctcc	1500
		cagctccttg				1560
		cagtgtgccg				1620
		ggctcgcacg				1680
gcagaagaag	atgcaggcag	ctgagttgtt	gtattctgat	aagagtcaga	ggtaactccc	1740
gttgcggtgc	tgttaacggt	ggagggcagt	gtagtctgag	cagtactcgt	tgctgccgcg	1800
		ctgacagact				1860
		ctgctgtgcc				1920
		ccctggaagg				1980
tgaggaaatt	gcatcgcatt	gtctgagtag	gtgtcattct	attctggggg	gtggggtggg	2040
gcaggacagc	aagggggagg	attgggaaga	caatagcagg	catgctgggg	atgcggtggg	2100
		ggtgctgaag				2160
		ctgtgacaca				2220
		atagctcagg				2280
					gcctccaaga	2340
gtgggaagaa	attaaagcaa	gataggctat	taagtgcaga	gggagagaaa	atgcctccaa	2400
catgtgagga	agtaatgaga	gaaatcatag	aatttcttcc	acttecteae	tcactgactc	2460
		tgcggcgagc		-		2520
						2580
		ataacgcagg				
		ccgcgttgct				2640
cgagcatcac	aaaaatcgac	gctcaagtca	gaggtggcga	aacccgacag	gactataaag	2700
ataccaggcg	tttccccctq	gaagctccct	catacactct	cctgttccga	ccctgccgct	2760
		ttctcccttc				2820
						2880
		tgtaggtcgt				
		gcgccttatc				2940
aagacacgac	ttatcgccac	tggcagcagc	cactggtaac	aggattagca	gagcgaggta	3000
tgtaggcggt	gctacagagt	tcttgaagtg	gtggcctaac	tacggctaca	ctagaagaac	3060
		tgctgaagcc				3120
		ccgctggtag				3180
						3240
		ctcaagaaga				
tcagtggaac	gaaaactcac	gttaagggat	tttggtcatg	agattatcaa	aaaggatctt	3300
cacctagatc	cttttaaatt	aaaaatgaag	ttttaaatca	atctaaagta	tatatgagta	3360
aacttggtct	gacagttacc	aatgcttaat	caqtqaqqca	cctatctcag	cgatctgtct	3420
		cctgactcgg				3480
						3540
		accaggeerg				
		ctttgttgta				3600
ctttgccacg	gaacggtctg	cgttgtcggg	aagatgcgtg	atctgatcct	tcaactcagc	3660
aaaagttcga	tttattcaac	aaagccgccg	tcccgtcaag	tcagcgtaat	gctctgccag	3720
		ttctgattag				3780
aatttattca	tatcaggett	atcaatacca	tatttttraa	agaggggttt	ctotastos	3840
		gttccatagg				3900
		acaacctatt				3960
agtgagaaat	caccatgagt	gacgactgaa	tccggtgaga	atggcaaaag	cttatgcatt	4020
		aggccagcca				4080
		tgattgcgcc				4140
uaaaccyt	Laccoactey	-Jac - Grace	-gag-gagac	sacacacycy	accyclylla	4740

```
aaaggacaat tacaaacagg aatcgaatgc aaccggcgca ggaacactgc cagcgcatca
                                                                               4260
acaatatttt cacctgaatc aggatattct tctaatacct ggaatgctgt tttcccgggg
atcgcagtgg tgagtaacca tgcatcatca ggagtacgga taaaatgctt gatggtcgga agaggcataa attccgtcag ccagtttagt ctgaccatct catctgtaac atcattggca
                                                                               4320
                                                                               4380
                                                                               4440
acgctacctt tgccatgttt cagaaacaac tctggcgcat cgggcttccc atacaatcga
                                                                               4500
tagattgtcg cacctgattg cccgacatta tcgcgagccc atttataccc atataaatca
gcatccatgt tggaatttaa tcgcggcctc gagcaagacg tttcccgttg aatatggctc
                                                                               4560 .
ataacacccc ttgtattact gtttatgtaa gcagacagtt ttattgttca tgatgatata tttttatctt gtgcaatgta acatcagaga ttttgagaca caacgtggct ttccccccc
                                                                               4620
                                                                               4680
ccccattatt gaagcattta tcagggttat tgtctcatga gcggatacat atttgaatgt
                                                                               4740
atttagaaaa ataaacaaat aggggttccg cgcacatttc cccgaaaagt gccacctgac
                                                                               4800
gtctaagaaa ccattattat catgacatta acctataaaa ataggcgtat cacgaggccc
                                                                               4860
                                                                               4867
tttcgtc
<210> 17
<211> 75
<212> DNA
<213> Artificial Sequence
<223> oligonucleotide
gatcaccatg gatgcaatga agagagggct ctgtgtgctg ctgctgtgtg gagcagtctt
                                                                                  60
cgtttcgccc agcga
<210> 18
<211> 78
<212> DNA
<213> Artificial Sequence
<220>
<223> oligonucleotide
gatctcgctg ggcgaaacga agactgctcc acacagcagc agcacacagc agagccctct
                                                                                · 60 ·
cttcattgca tccatggt
                                                                                 78 .
<210> 19
<211> 33
<212> DNA
<213> Artificial Sequence
<223> oligonucleotide
<400> 19
ggtacaaata ttggctattg gccattgcat acg
                                                                                 33
<210> 20
<211> 36
<212> DNA
<213> Artificial Sequence
<223> oligonucleotide
<400> 20
ccacatctcg aggaaccggg tcaattcttc agcacc
                                                                                 36
<211> 38
```

<212> DNA <213> Artificial Sequence	
<220> <223> oligonucleotide	
<400> 21 ggtacagata tcggaaagcc acgttgtgtc tcaaaatc	.38
<210> 22 <211> 36 <212> DNA <213> Artificial Sequence	
<220> <223> oligonucleotide	
<400> 22 cacatggatc cgtaatgctc tgccagtgtt acaacc	. 36
<210> 23 <211> 39 <212> DNA <213> Artificial Sequence	
<220> <223> oligonucleotide	
<400> 23 ggtacatgat cacgtagaaa agatcaaagg atcttcttg	. 39
<210> 24 <211> 35 <212> DNA <213> Artificial Sequence	
<220> <223> oligonucleotide	
<400> 24 ccacatgtcg acccgtaaaa aggccgcgtt gctgg	35
<210> 25 <211> 4864 <212> DNA <213> E. coli (V1R plasmid)	
<400> 25 tegegegettt eggtgatgae ggtgaaaace tetgacacat geageteeeg gageagettgtet gtaageggat geegggagea gacaageeeg teagggegeg tea	
ttggcggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctg accatatgcg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atc ctattggcca ttgcatacgt tgtatccata tcataatatg tacatttata ttg tccaacatta ccgccatgtt gacattgatt attgactagt tattaatagt aat	gagagtgc 180 cagattgg 240 ggctcatg 300
ggggtcatta gticatagce catatatgga gttccgcgtt acataactta cgccccgcctggc tgaccgcca acgacccccg cccattgacg tcaataatga cgtcatagtaacg ccaataggga ctttccattg acgtcaatgg gtggagtatt tactgccacttg gcagtacatc aagtgtatca tatgccaagt acgcccccta ttg	tatgttcc 480 cggtaaac 540
tgacggtaaa tggcccgcct ggcattatgc ccagtacatg accttatggg actttggcagtac atctacgtat tagtcatcgc tattaccatg gtgatgcggt tttcaatggg cgtggatagc ggtttgactc acggggattt ccaagtctcc accgtcaatggg agtttgttt gccacaaaa tcaacgggac tttccaaaat gtc	tttcctac 660 tggcagta 720 cccattga 780

actocytita gigaaccyte qaategecyte gagacycat cacgetgit tyacetcea 960 tagaagacae egggaccyat caagecteeg eggecyggaa eggtycatty gaacgeggat 1020 tectatagea getatacteeg tittitggett ggggtetatt accectagge 1080 tectatagea gitaagatta gectataget tataget ataggeca eccectagge 1180 tataggtgat gitaagetta gectataget gggggtatatt gacattatt gacactacte 1200 ctattggtga egatactte cattactaat cataacatg getettge acaactatet gattittea 1120 aggatggggt etcattatt attacaata teacataacat getettge acaactact 1260 ctattggtga egatactte eatacaata cacacacage getettge acaactact 1260 ctattggtga tatagetat eatacaata cacacacage accacacaca cacacacaca gagatggggt etcataacat acacacacac gagacacacaca cacagtggac tectacacacacacacacacacacacacacacacacacac							
agactogitta gipaaccgic agategocig gagacgocat caaccgitt tigaactgocat 19aagacgocgat caccogtgoctococgigocat gigaacgocgat caccogtgocatcgic gaacgocgat caccocgitoc tectocatgit 1140 ataggitadig statagotta gittagotta caccocgit tectocatgit 1140 ataggitadig statagotta gittagotta caccoccitic sectocatgit 1140 ataggitadig statagotta gittagotta caccocgit tectocatgit 1140 ataggitadig statagotta cactoficcit cagagactga cagactata acaccacci tectocatgit 1140 ataggitadig statagotta cactoficcit cagagactga cagagattagi citatitata atticacata cacacataca acaccacci teccocatgoci cagagattgit tagticata atticacatac cacacacacacac eccatycocci caccagitit tattaacat acactgoggat citacataca acaccaccac eccatycocci acaccagtic tectocogga gogocgac tectocatacc agaccotycic cocatycocci cagagacta tagticocaca cagactytic cacacacacacacacacacacacacacacacacacac	ctccgcccca	ttgacgcaaa	tgggcggtag	gcgtgtacgg	tgggaggtct	atataagcag	
tagaagacaa cyggaccyat cagactecy cygcegygaa cygtycatty gaaccycgyat 1020 ttettatyca tyctatacty tttttygcty gygyttatt accecctyct tectatytt 1140 atagytyatg ytatagytta getatacytt getatacyg tyggyttatt acceccyat tectatytt 1140 atagytyatg ytatagytta getatacytt cygagactya cagactacty tydyttycyatty gatacytty catacygyatty gygyttycyatyatyatyatyatyatyatyatyatyatyatyatyaty	agetegttta	gtgaaccgtc	agategeetg	gagacgccat	ccacactatt	ttgacctcca	960
tecteatgree aagagtgaeg taagtacege ctatagagte tataggecea cececttgge 1080 tectatatgra tgetatategt tettitgett ggggtetata accocceget tecteatgtt 1140 ataggtgatg gtatagetta gectataggt gtgggttatt gaccattatt gaccactect 1200 teattggta gtatagetta gectataggt gtgggtatt gaccattatt gaccactect 1260 teattggta tatgecaat cactycett cagagactga cacgagactg statttata 1220 aggatggggt tetteattatat tatecaaat teacactata acceggaattet tattacaaat tecacgataa cacgagactga cacgagactga cacgagactga tettetaggta gtgtttetagg aggatggggg tetteacgaga tetteaggta gtgtttetagg accacacaca cacgagtgggg tetteacacgaga tetteacacac gaggagaga gtgttgcac gagagagagagagagagagagagagagagag ggggttgcac gacagagagagagagagagagagagagagagagag	**********	gagagagat	agaregees	222222222	caatacetta	gaacgcggat	1020
tetetatgota tgotatacty tttttgott gggstetta cacccccgst tectcatgtt 1200 ctattggtga gatactts cattactaat cattactaat cattactaat tagacattatt gaccattatt 1320 aggatggggt ctcatttatt atttacaaat cacacaatg gctctttgcc acaacttct 1320 aggatggggt ctcatttatt atttacaaat teacaataac acacgaaccts gtatttttac 1320 aggatggggt ctcatttatt atttacaaat teacaataac acacgaaccts gtatttttac 1320 cacacacac typtcoctog gagctcctt gccacagagg ctcaccacacacacacacacacacacacacacacacaca	Layaayacac	cygyaccyac	ccaycccccy	cygccygaa	cggcgcaccg	gaacgcggac	
tatagtgatg glatagetta getatagett getataget grigggttatt gaccattatt gaccactecc tlatitiggta egatacttic cattacata cateacata getetitiges cacaactect 1260 tatitiggta egatacttic cattacata cactegora cacaactect 1260 tatitiggta tatgecaata cactegorate cacagaget cacagaget cacagactet gratititate 1380 cegragitit tattacacat acetgoret cacagagatg cacagagatget tectacate acetgoret cacagaget 1380 cegragitit tattacacat acetgoret tectacagega atotogggta cytythoogg 1480 cagocacta tygitogetog geogrogoret tectacate gagocetyst cocatgore 1500 cagocacgatg cocaccacca cegitytoca cycitytocac groupers grou	tccccgtgcc	aagagtgacg	taagtaccgc	ctatagagtc	tataggccca	ceeeettgge	
ataggtgaty gataagctta gcctataggt gtgggttatt gaccattatt gaccaatccc 1260 ttattggcta cagatatttc cattataat cataacatg gactetttyca caacatctct 1260 ttattggcta tatgccaata cactgtcctt cagagactga cacagactct gtatttttac 1380 cagagatggg ctcatttatt attataaat cacatatac aacacacac	ttcttatqca	tgctatactg	tttttggctt	ggggtctata	cacccccgct	tcctcatgtt	1140
tattggtga cyatactttc attactaat cattaacatg getetttgec acaactetet 1320 aggatggggt cteatttat attacaaat teacataata cacagactet gfattttta 1320 aggatggggt cteatttat attacaaat teacataata aacacacacg teeccagtge 1340 cacaggactt teattaacat aactgggat teaccaggac teaccagggact gagtagggga gagtagggg 1440 acatgggete tettecggta geggggggg tetetaaca gagcectget cecatgect cacagggacta tettecagga gagtagggact gagtagggac gagtagggac gagacaggacgacacacacacacacacacacac	atammetm	gtatagetta	acctataggt	gtgggttatt	gaccattatt	gaccactccc	1200
tetattggeta tatgecaata cactgteett cagagactga cacggactet gtatttttac 1380 cagaatgtgggt cteatttat attataaaat teacaataca aacacacacg tecccagtget 1380 cagaagtttt tattaaacat aacgtgggat cteacatget 1500 cagcagtttt tattaaacat aacgtgggat cteacatget 1500 cagcagattt tattaaacat aacgtgggat cteacatget 1500 cagcagatte teteceggta geggegggget tetecaacg gagcettget cecatgete 1500 cagcacgatg cecacacacc cagtgtgcc geacaagge gtygagggag ggstatgtac 1620 tyaaaatgag ctoggggag gggttgcac geacaagge gtygaggag gtyaaggtag 1620 tyaaaatgag ctogggggg gggttgcac gecaaagge gtygaggag tttgagacag tttaggaaga ttaaggagac 1800 gecagcacac agacataata getyaaggaa tyaaggtet tydgtettga taaggactag ttgtgttege ctgcacacacaca cagtgtggg tyaagtetg ttgtgttetga taaggactg ttgtgttege ctgcacacacacacacacacacacacacacacacacacac	-t-t		gcccacaggc	9099900000		acanatatat	
aggatggggt ctcattata atttacaaat tcacatata acaccaccg tccccagtgc 1340 cacagggatt ttattaaacat acgtgggst ctcacaggga tcccacggg 1440 cacagggatt ttattaaacat acgtgggst ctccacagga attcagggat 2 ggggggggg gggtcctt gccataggtgcc 2 ggacctgggag 2 ggcccacaca cagtgtgcc gcacaggacg gggcggag gggtgggggggggg	ctattggtga	cgatactttc	Callactaat	ccataacatg	getettiget	acaactccc	
aggatggggt ctcattata atttacaaat tcacatata acaccaccg tccccagtgc 1340 cacagggatt ttattaaacat acgtgggst ctcacaggga tcccacggg 1440 cacagggatt ttattaaacat acgtgggst ctccacagga attcagggat 2 ggggggggg gggtcctt gccataggtgcc 2 ggacctgggag 2 ggcccacaca cagtgtgcc gcacaggacg gggcggag gggtgggggggggg	ttattggcta	tatgccaata	cactgtcctt	cagagactga	cacggactct	gtatttttac	
cogcagttitt tattaaacat aacgtgggag ctccacgag atctcgggt cgtgttcgg 1440 caatgggact titctcagtga gcaggagact titctacatca gagcctgtc cocatgcctc 1500 cagcagactca tigtgcgctg gcagcactct gctcacaca gagcagactcac 1560 cagcagactac tigtgtcgctg gcagcactcct gcacaagac gtggggggag gactaagga 1560 caagaagag cacacacaca cagttgtgc gcacaagac gtgggggtag ggtatgtac 1560 ggcagaagaa gatgaggac ggtggtgta cacaagacc gtgggggtag ggtatgtac 1680 ggcagaagaa gatgaagac gctagattgt tytgttctga taaagatcac gagtactcc 1740 gctgcgccac agacataata gctgacagaa taacagactg tictttcaca tgggtcttt gctttaacgg gtgacgaaa tacagactg tigtagtcga gcagcactcc ctccccccgt gcttcttga tectgagtag gtgcactcc cactgtcctt tectaataa 1980 atagagaaat tgcatcgcat tytcgagta ggtgtcattc tattctgggg ggtgggtgg 2040 ggcacacac gccttccttg accetggaaga gaccagacacc ctcctctggg agacagacacc ctctctctgt gacacaccc tyccacgcc ctccctgggc gagaagaaa tgagagaagaa tatgagagaa tccacacacc tyccacgcc ctcgcttct attcaggag cactcaccactaatag acactcataag cacacaccc tyccacgcc ctcgcttct attccagag accactcacatag cacacacacc ctcatcagacg cacacacacacacacacacacacacacacacac	aggatggggt	ctcatttatt	atttacaaat	tcacatatac	aacaccaccg	tccccagtgc	1380
acatgggetc tetecagaa geggeggag tetetacatac gagcoctgot cocatgotic cagcagatca tygtogotog gagstettt getectacaaa gtggaggca gactaggca 1560 cagcagatac tygtogotog gagstettett getectacaaa gtggaggca gactaggca 1560 gagaaaatga cocacacca cagtgtgce gacaaaggc gtggoggtaa gattatgtt 1620 tgaaaatgag cocacacaca gagcatacac geggeggaaa gagaaagaa gagcaaggca gtgagttgca coctacaccac agacataca gtggaggaa gtgagtagta tttgttetcaa taaaggaact taaaggaact gtgagggcaa gtgagacaa gagtaactcc 1740 geggegcaca agacaaataa getgacagaa taaacaagaact tecetttcaa tgggtetttt 1860 ctgcagtcac cytecttaga tetyctgtgc ettetagtg cocacacca catggetett tecttaaaa 1980 ataaggaaat tgaategaa tetyctgtgc gtgcagcac cactgfectt tectaaaaa 1980 ataaggaaat tgaategaa tgtggaagaa gagacactcc cactgfectt tectaaaaa 1980 aggcacacac cacaggggag gattgggag gagcagacac cacaggggag gagaaaatgg gacacacca cacaggggag gagaaaatgg gacacaccac gtcacaggag gtggggggg gagaaaacac cettectetg gacacaccc gtcacacacc gtcacacacc gtcacacacc gtcacacacc gtcacacacc tggggaggaggagaaaa taagaagaaa gacacaccc gtcacacacc cacagaggaga ctcacacacc ctcacagagagaa aaagaaagaa gacattataag teacaagaga gaaaaatgc ctcacacac gtcacaagaac aaggaaagaa gagaaaatt tettecgett ccccgctogt tetteggetgg gagagggta taagcaggaa gagaaaatgc ctcacacagagaacaggaagaaaacaggacagg							1440
cagogactea tygtegeteg geageteett getectaaca gtggaggec gacttagga 1560 cagoacyaty cccaccaca cagtytyce geacaaggec gtggggtyag gytatytyc 1520 tygaaatyag ctcggggag gyggttygae cgctyaacyaa tttggaagaa ttaaggcag 1680 gygagaagaa gatgcaggea gytgagtyt tytyttetga taagagtaagtea gytaactec gytgggggag ctgytaacya taataag gtgaactea tttggaagaag taggtaactea gytgacagaa taacagact ttccttacta tygytttyge 1740 cygggagcaay tactgyaagaag tygacaactea ttctoatyag ttgccaccac agacataata getyacaagaa tygacaactea catcgatea gytgccactee cactgtcctt tytyttyge 1920 cetececcyf geetteettag acctggaag gytgccactee cactgtcctt tectaataaa 1880 atagaggaaat tygacaactea catcgagaag gytgccactee cactgtcctt tectaataaa 1880 agagcacaate cactgagaag gattyggaaa gattagagg gytgggggggggg							
cagcagatg cccaccacca ccagtgtgcc gcacaaggcc gtggcggtag ggtatgtgc gaaaaggc gcagaagga gtgaagtcgaaggc gctgaagtcg tgtgtaccac gctgaagacactttgaaagacgac taagagaagaa gatgcaagac gctgagttgt tgtgttctga taagagtaag aggtaactcc gttgctggtg ctgttaacgg tgaagacat taagagacag taagatgaagac gtaagacaccac agacataata gctgacagac taacagactg ttcctttcac tgggtctttt 1960 ctgcagtcac cqtccttaaga tetgctgtgc cttctagtg ccaccac cactgtcctt tcctaaaaa 1980 ataaggaaat tgatcccagt gccttcccttg accttggaag gtgccactcc cactgtcctt tcctaataaa 1980 ggcagcacac cacaggaggag attgggaag acatatgaag gcatgctggg gatgggggg ggaggaag acagagaagacaccaccactgtccttggaag gtgcagcaccac cctgttcttt gttccaagag ggcagcaacac cacaggtgc ttgaagaatt gaccaggtc ctcctctggaag gtgcagcaccac ctgtttctta gttccaggcc 2220 cactcatagg acactcatag ctcaaggaaga ctcagagaga ctcgactct aatccacacc gctaaagaact ccttagagaagaac tcaagaagag ctcagagaagaag acactcataga cacaagaagaag acacaacaca	acatgggctc	ttctccggta	gcggcggagc	ttctacatcc	gagecetget	cccatgcctc	
cagcagatg cccaccacca ccagtgtgcc gcacaaggcc gtggcggtag ggtatgtgc gaaaaggc gcagaagga gtgaagtcgaaggc gctgaagtcg tgtgtaccac gctgaagacactttgaaagacgac taagagaagaa gatgcaagac gctgagttgt tgtgttctga taagagtaag aggtaactcc gttgctggtg ctgttaacgg tgaagacat taagagacag taagatgaagac gtaagacaccac agacataata gctgacagac taacagactg ttcctttcac tgggtctttt 1960 ctgcagtcac cqtccttaaga tetgctgtgc cttctagtg ccaccac cactgtcctt tcctaaaaa 1980 ataaggaaat tgatcccagt gccttcccttg accttggaag gtgccactcc cactgtcctt tcctaataaa 1980 ggcagcacac cacaggaggag attgggaag acatatgaag gcatgctggg gatgggggg ggaggaag acagagaagacaccaccactgtccttggaag gtgcagcaccac cctgttcttt gttccaagag ggcagcaacac cacaggtgc ttgaagaatt gaccaggtc ctcctctggaag gtgcagcaccac ctgtttctta gttccaggcc 2220 cactcatagg acactcatag ctcaaggaaga ctcagagaga ctcgactct aatccacacc gctaaagaact ccttagagaagaac tcaagaagag ctcagagaagaag acactcataga cacaagaagaag acacaacaca	cagcgactca	tggtcgctcg	gcagctcctt	gctcctaaca	gtggaggcca	gacttaggca	
tgaaaatgag ctcgggagc gggttgac ggcttgac ggctgaca tttggaagac ttaaggcagc 1680 ggcagaagaa gatgcaggca gctgagttgt tgtgttctga taagagtcag aggtaactcc 1740 cgttgggtg ctgttaacgg tggagggggg tgtagtctga gagtaactc tgtgcgcgcacc agacataata gctgacagac taacagactg ttccttcca tgggtcttt 1860 tcgcagtcac cgtccttaga tetgctgtgc cttctagttg cagcactc tgtgttttcc 1920 cctccccgt gccttccttg accctggaag gtgccactc cactgtcctt tcctaataaa 1980 ggcgacaac caacagaact ttcttccac tgggtctttt 2920 cctcccccgt gccttccttg accctggaag gtgccactc cactgtcctt tcctaataaa 1980 gcctatatagg taccagat tgtgagaaga gatgggaga gatggggag gatgggggg gtgggggggg	cagcacgatg	cccaccacca	ccagtgtgcc	gcacaaggcc	gtggcggtag	ggtatgtgtc	1620
ggcagaagaa gatgcaggaa gctgagttyt tytyttctga taagagtcag aggtaactcc 1740 ggttgcggt cytytaacgg tygaggcag gttagtctga gcagtactcg tyctyccgc 1800 gegcgcacc agacataata gctgaagac taacagactg tectteca tyggtettt 1860 ctgcagtcac cytecttaga tetgetgtge ettetagtg cagacact tyttgtcc 1920 ceteccecg gcettecttg accetygaag gtgcactce cactytect tectaataaa 1980 ggcagcaca caagggggag gattgggaag acaatagcag gcatyctggg gatgggtgg 2040 gctctatggg taccaagggag gtgtggaag acaatagcag gcatyctggg gatgggtgg 2100 gctctatggg taccaagggg cytaagaatt gacacacct gtcacagce ctggttett gttcagccc 2220 cactcatagg acactcatag ctcaggaggg cytaacaccc tytacaggac gataggagg cytacatca aggagaaga atgagaagaa gcacaccat gtcaggaggg ctctcccc ctcatcagc caccaacaa aacctagcc ctagaagaag ttggagaggg ctctcccc ctcatcagc caccaaacaa aacctagcc caagaagagg tgaggaagat atgagaagaa ggcatataag tycagagggg gagaaaatg ctcaagagg tgaggaagta atgagaagaa ggcagtgag tacagcacct gagttegg tgaggaagta atgagagaaa tcatagaag tycaacacct caaaggggg aacaaggacg tcacagaa cagggggaag agcaagag acactagcc ctcgctcac tyaactcgctg tgaggagtgt tttggctggg ggaggggggggggggggg	tassastasa	ctcaaaaaaa	acacttacec	cactaecace	tttqqaaqac	ttaagggagg	
ggtgtgcaca agacataatag gtgaagggaag gtaagtctga gcagtactcg ttgttgtgctttt 1860 ctgcagtcaca cagtacttaga tctgttgtgc cttctagttg ccagtcacac gtgtgtttgtg 1980 cttctcagtgaagaat tgcattcagt gttgttgtgc cttctagtg gtgcacacc gatagtgggaagaat tgcatcagaaga gtgcacacc cactactagtg tacctttgg cacatggggaag gattgggaag gtgggtgtcatttgg taccaaggggaag gattgggaag acaatagaag gatgtgggg gatgcggtgg 2040 ggcacacac ccttctctgt gacacaccct gtcagtacc ctctctggg cacatagaagaag cacatagaagaagac caccactataga ctcaagaagag gtcagtggg gtcagtggggggggacactc cttctctgt gacacaccct gtcaagacac cccctactaag ctcaataga ctcaagaagaagac tcagcctta aatccaacc gtcaaagtac 2280 gaagaaata aagaagaa gtcataag cacaaacca aactagcct cacaagatg 2280 tagagaagaa tagagagaa tcataagaat tcttccgtt cacacaccac gtaaagtac 2280 gaagaaagaa atgagagaa tcaatagaat tcttccgtt cctcgctcc ctaacagagg 2340 tagagaacac aaaaggggaaa ccaagaagaagaa acaatgaagag gaaaaaagc ctcaaacaca 2400 tagagaagaa atgagagaagaa acaatgaagag gagaaaagc caaaagggcg 2580 agaaaccgta aaaaggccg gtgstgggg tttttccata ggctcagcc cacaagaggg 2580 cacacaaaa atgagagaag ctccctctggg ggtgtgggg tttttccatag ggctcgcgcc ccctgagaag 260 caccacaaa atgagaaga ctccctcgtg ggctcacct caaaggggat aaaaggac 2700 caggcgtttc ccctggaag ctccctcgtg ggctctcctg ttccgagcac gcaaaaggc 2700 caggcgtttc ccctggaag ctccctctgt ggctctcctg ttccgagcac gcaaaaggc 2700 caggcgtttc ccctggaag ctcactcac gacgggggggggg	tyaaaatyay	cccggggagc	gggcccgcac	tetegacyca	teeggaagae	ccaaggcagc	
gegegecace agacataata getgacagac taacagactg theothea tigglethte chiquagheac eghechtaga tedgetgaga gtgecachee cachgheath thechaataaa atgagagaat tgategaag gtgecachee cachgheath thechaataaa agagagaaat tgategaag gtgecachee cachgheath thechaataaa 1980 agagacacag caaggaggag gathgygaag acaatagaag gcatgetggg gdgggggggg gedgggagacacag caaggaggag chacagagta gacacacct gtccacgace ctcetggge agaaagaaga 2160 agagaaatac cactactetgg ctcaagagaat gacacggtee ctcetggge agaaagaaga 2220 cactaatagg acaactaatag ctcaagaggg ctcacacaca aactaagact gdtaaagaagaagaagaagaagaagaagaagaagaagaagaa	ggcagaagaa	gatgcaggca	gctgagttgt	tgtgttctga	taagagtcag	aggtaactcc	
gegegecace agacataata getgacagac taacagactg theothea tigglethte chiquagheac eghechtaga tedgetgaga gtgecachee cachgheath thechaataaa atgagagaat tgategaag gtgecachee cachgheath thechaataaa agagagaaat tgategaag gtgecachee cachgheath thechaataaa 1980 agagacacag caaggaggag gathgygaag acaatagaag gcatgetggg gdgggggggg gedgggagacacag caaggaggag chacagagta gacacacct gtccacgace ctcetggge agaaagaaga 2160 agagaaatac cactactetgg ctcaagagaat gacacggtee ctcetggge agaaagaaga 2220 cactaatagg acaactaatag ctcaagaggg ctcacacaca aactaagact gdtaaagaagaagaagaagaagaagaagaagaagaagaagaa	cattgcggtg	ctgttaacgg	tggagggcag	tgtagtctga	gcagtactcg	ttgctgccgc	1800
ctgcagtcac cgtccttaga tctgctgc cttctagttg cagccatt gttgttgcc 19320 cctcccccgt gccttccttg acctgagag gtgccatcc cactgtcctt tctataaaa atgaggaaat tgcatcgat tgtctgagta ggtgcattc tattctgggg ggtgggtgg 2040 ggcagcacag caagggggag gattgggaag acaatagcag gcatgctggg gatggggtgg 2160 aggcacatc ccttctctgt gacacacct gtccacgcc ctggttctta gttccagccc 2220 cactcatagg cactcatag ctcaggaggat gaccaggtct ctcctgggcc agaaagaacg ttggagcggt ctctccctcc ctcaggagggg gtccgcctt aatccaccc gtcaaagtac gagaaatta aagcaagata ggctattaag tgcagagggg gagaaaatgc ctcaacaca tagagagat atgagagaaa tcatagaatt tcttccgtt cctcgctca ctcaacaca gcgtcggtgg ttcggctgcg gcagagggata tcagctcact caaaggcgg aaaaaggcc gcacacacaaa accaaaaa accaacaaa accaaaaca aacaaaaa acgacgcc gagaggagaaaa gacagtggag gagaaaatgc ctcaacacag aggaaccgta aaaaggccgc gttgctggcg tttttccata ggctcacc ccctgacgag aggaaccgta aaaaggccgc gttgctggcg tttttccata ggctcaccc ccctgacgag aggatacctg ccccttggag ctccctctgt gcgtctcctg ttccgaccc ccctgacgag aggatacctg cccctttct cccttcggga agcgtgaga gcggtgagga ccggcttcc gagtacctc ccctggggaggta ccgccttctt cccttcggga agcgtgggg ttttccaaggcg tttccaacaaa atcgacgcc cccttcggga gtcgttcgc tccaagctgg gctgttccagcc ccctgacgag aggtatctc cccctggaag ctccctcgt gcgttctcctg ttccaacaca gcgggtacacacacacacacacacacacacacacacacac	gcgcgccacc	agacataata	gctgacagac	taacagactg	ttcctttcca	tagatetttt	1860
ceteccegt gettecting accinguag grigecation cacinguated type to the state of the sta	ctacaatcac	catactteas	tetactatae	cttctagttg	ccacccatct	attattacc	1920
atgaggaaat tgcatcgcat tgtctgagta ggtgtcattc tattctgggg ggtggggggg ggcggcacacacaca caagggggag gattgggaga acaataacaca gcatctgtgg gattggggggg 2100 gctctatggg tacccaggtg cttgaagaatt gacccggttc ctcctgggcc agaaagaagc 2160 aggcacatcc ccttcttctgt gacacaccct gtccacgccc ctggttctta gttccagccc 2220 cactcataag acactcatag ctcaggagggggggggg	Ctycayttac	cyccccaga	cetgetge	ccccagccg	b-b-	besteere	
gcctacacaa caaggaggag gattgggaag acaatagcag gcatgctggg gatcggtgg 2160 aggcacatcc ccttctctgt gacacaccct gtccacgccc ctggttctta gttccagccc cactactagg acactcatag ctcaggaggg ctcccctcc ctcatcacgcc acactcactagg trygagggg ctccctcc ctcatcagcc caccaacca acactagcct ccaaggaggg trygaggagt ctccctcc ctcatcagcc caccaacca acctagcct ccaagagtgg gagaaatta aagcaagata ggctattaag tgcagaggga gagaaaatgc ctcaagagtgg tcacacagaa tcatagaa tcatagaat tcttccgctt cctcctcc tcaacaggc tcacacagaa tcaggggataa cgcaggaaag aacatgtgag caaggaacgt aaaggccg gttyctgggg tttttccata ggctcgccc cctgagagg caaggaggta atgaggaaa tcatagaagt tcagctcact caaaggcgg tcacacagaa caggggataa cgcaggaaag aacatgtgag caaggagcta aatacggccg gttyctgggg tttttcacta ggctccgccc cctgagaag caggacgttc cccctggaag ctccctcgtg cgctcctctg tcccacagaac caggggtttc cccctggaag ctccctcgtg cgctcctctg ttccgacct gccgttacc ggatacctg gttggggag ggtggttgc tccaaggtgg gcttgttgga gtcgttacc ggttcagccg accgttgcc cttatccgga agcgtggcgc tttctcaatg gttcaggcag accgttgcg cttatccgg actaactacg gctgagagag aacaggagagagagac cacgacttat cgccactggc agcagccact ggtaacagga ttagagaagag gaggatatga ggcggtgctc cagagttct ggaaggcagt gactacacag gctacactag ggcggtgcta cagagttct ggaagcagt gctaactacg gctacactag agagatagta cagactatat cgccactggc gacagcact ggtaacagga ttagagaag cgcggaaaaa aagacaccgc tggtagggg cttaactagg gctacactag cgcggaaaaa aagaccccc ttagtgggg cctaactacg gctacactag cgcggaaaaa aagaccaccg tggtaggggg gctttatttt gagacaaa cacaccgc tggtagggggggggggggggggggggggggg	cctcccccgt	gccttccttg	accctggaag	gtgccactcc	cactgtcctt	teetaataaa	
gcctacacaa caaggaggag gattgggaag acaatagcag gcatgctggg gatcggtgg 2160 aggcacatcc ccttctctgt gacacaccct gtccacgccc ctggttctta gttccagccc cactactagg acactcatag ctcaggaggg ctcccctcc ctcatcacgcc acactcactagg trygagggg ctccctcc ctcatcagcc caccaacca acactagcct ccaaggaggg trygaggagt ctccctcc ctcatcagcc caccaacca acctagcct ccaagagtgg gagaaatta aagcaagata ggctattaag tgcagaggga gagaaaatgc ctcaagagtgg tcacacagaa tcatagaa tcatagaat tcttccgctt cctcctcc tcaacaggc tcacacagaa tcaggggataa cgcaggaaag aacatgtgag caaggaacgt aaaggccg gttyctgggg tttttccata ggctcgccc cctgagagg caaggaggta atgaggaaa tcatagaagt tcagctcact caaaggcgg tcacacagaa caggggataa cgcaggaaag aacatgtgag caaggagcta aatacggccg gttyctgggg tttttcacta ggctccgccc cctgagaag caggacgttc cccctggaag ctccctcgtg cgctcctctg tcccacagaac caggggtttc cccctggaag ctccctcgtg cgctcctctg ttccgacct gccgttacc ggatacctg gttggggag ggtggttgc tccaaggtgg gcttgttgga gtcgttacc ggttcagccg accgttgcc cttatccgga agcgtggcgc tttctcaatg gttcaggcag accgttgcg cttatccgg actaactacg gctgagagag aacaggagagagagac cacgacttat cgccactggc agcagccact ggtaacagga ttagagaagag gaggatatga ggcggtgctc cagagttct ggaaggcagt gactacacag gctacactag ggcggtgcta cagagttct ggaagcagt gctaactacg gctacactag agagatagta cagactatat cgccactggc gacagcact ggtaacagga ttagagaag cgcggaaaaa aagacaccgc tggtagggg cttaactagg gctacactag cgcggaaaaa aagaccccc ttagtgggg cctaactacg gctacactag cgcggaaaaa aagaccaccg tggtaggggg gctttatttt gagacaaa cacaccgc tggtagggggggggggggggggggggggggg	atgaggaaat	tgcatcgcat	tgtctgagta	ggtgtcattc	tattctgggg	ggtggggtgg	2040
getettatggg tacceaggtg ctgaagaatt gaccggttc ctcctggcc agaaagaaga 22160 aggacacatc cettetetg gacacacct gtccacgcc ctggttetta gttccagcc 2220 cactcatagg acactcatag ctcaggaggg ctccgccttc aatccacc gctaaagtac 2280 ttggaggggg ctccccccc ctcatcagcc cacaaacca aacctagcc ccaagagtg 2340 gaagaaatt aagaagaata ggctattaag tgcaagaggg agaaaatgc ctccaacatg 2400 ttgaggaagta atgagagaaa tcatagaatt tcttccgctt cctactcctc gctaacagag 2460 cgctcggtcg ttcggctggg gggagggggta tcggcaggaa aacatgtgag caaaaggcc 2520 tccacagaat caaggggataa cgcaggaaaag aacatgtgag caaaaggcc gcaaaaaggcc 2520 aggaaccta aacaggccg gttyctgggg tttttccata ggctccccc cctgacagga 2520 cacacaaca atcgacgcc aagtcagagg tggcgaaacc cgacaggacac cccctggaag ctccctcgtg cgctctcctg ttccgacct gccgcttacc ggatacctca gtcagctca cccctgggaag ctccctccgtg cgctctcctg ttccgacct gccgcttacc ggatacctca gtcaggtag ctccctcggg agcgtggcgc ttctccaacaacaaccgct cgcctttct cccttcggga ggcgggaacc cgaacaggac ttcggctgg cccacactaa cgcacctca gccgcttcc gacacctca ggcaggacgc cttatccaggat acctacaggataccggacggcggcgcactacacaggacggcgcc cacactaac ggcagcact ggaagcagata accacggacact ggaagcagata accacgacgacagaacact ggaagcagaaa accacgga ggaagcagata accacgacgacagaacact ggaagcagaacacacacgacactacag gctaacactag gctaacacacacacacacacacacacacacacacacacac	ggcagcacag	саадададад	gattgggaag	acaatagcag	acatactaga	gatocogtog	2100
aggcacatcc cettetetg gacacacect gtecacgee etggttetta gttecagee 2220 cacteatagg acatetatag gacatetatag getegeette aateccace getaagtae 2280 tuggaggggt etetecete etcateagee caccaaaca aacetageet ceaagagtgg 2340 gaagaaatta atgagagaaa gegatataag gegatateag gagaaaatge etetegetge tuteggetgee gegageggta teagetaget etegetgeet eteggetgee gegageggta teagetaget eaaggggata atgagagata etcatagaat tetteegett etegetgeet eteggagggata etggagaggat atgaggaggt etecacagaa etagggata etggagagaga aacatgtgag caaaaggee getaaggaggata atgaggatae etggagagggata teagetgag caaaaggeet aaaagggata atgaggatae etggagagaga aacatgtgag caaaaggeet getaacagag ggedaaacegta aaaaggeeg gttgetggeg ttttecata ggeeteegee ceetgaagag 2580 aggaataceta ategageete aggetggegg ttttecata ggeeteegee ceetgaagag 2580 aggatacetgt eeggettte eeggtgtagga gtegetgetget ggetgeteet gegettatee gegetteete eeggtgtagga gegtggggg etteteeteggag etgetgetgee tetegagagetgggetteegeetgeeggeetteetegggagggggggg	antatataaa	tacccacata	atanaanatt	asaccastta	ctcctcccc	242224224	
cactcatagg acactcatag ctcaggaggg ctccgcttc agagacggc ctctcctcc ccacaaacac aactcataggagggg ctctcacctcc ccacaaacacacacacacacacacaca							
ttggagcgt ctctcctcc ctcatcagc caccaaaca aacctagct ccaagatgg gagaaatta aagcaagta ggctattaag tgcagagag gagaaaatgc ctccaacatg 2400 gagaaaatta tcagagatt tctccgctt cctgctcac tgactcgtg 2460 cgctcgtgtg ttcggctgcg gcgagcggta tcagctcact caccaggaat ccagggaata ccagggaata ccagggaata ccagggaata cagggaata aagtcagagg ttgcggaaacc caaaggcgg taaacaggcc 2580 gagaacacta cccctggaag ctccctcgtg cgctccttg ttccaacata aagtcagagg ttgcggaaacc cgacaggact ataaagatac 2760 ggatacctgt cccctggaag ctccctcgtg cgctccttg ttccaacatg gccgcttacc 2760 ggatacctgt cccctggaag ctcacctagg agcagtactg acaaaggca gagaccccc 2760 ggatacctgt cccctggaag ctcacctagg agctgtgcgc tttctcaatg ctccacgtac 2760 ggatacctga gttggtgaa agcagtgcgg ttttctcaatg ctcacgctt 2820 aggtatctc ggtcgtccc ctatccgga agcagtgcgg ttttggagcg ctaacagga 2940 cacgacttat cgccactagc aagagcagt aacttcggta ttggagaag cggagaacacc 2880 ggcggtggtca cagagttct gagagcagt acatcagga gaggtatgt 3000 ggcggtggtca cagagatct tgaagcagt acatcaggaa 2940 ttggatact ggtaccacag aagagaccat ggttaccatagga gaggtatgta 3000 ggcagaaaaa aagaatctaa aggaatttt ggcagagaa aaaagagttg ttggagaaga 2940 ttggatact ggtaccacagaa aaccaccgc tggtagaggg ggtttttttg tttggaagaa aaccacaggaa acaggaagaa aagagtaca ccacgtaa aagaagtca 2760 ggtacaggaggaggggggggggggggggggggggggggg	aggcacatcc	ccttctctgt	gacacaccct	gtccacgccc	ctggttctta	gttccagccc	
gaagaaatta aagcaagata ggctattaag tgcagaggga gagaaaatgc ctccaacatg 2400 tgaggaagta atgagagaaa tcatagaatt tcttccgctt cgctcggtcg tcggctgggcg gcagacgggga tcagctcact caaaggcggta caaggcgga aaaaggccg gcagacgggg tcagccact cacacaaa atgagacgc ggtgctgggg tttttccata ggctccgccg caaaaggccg gcaacaaaa accacaaa atgagagg tgcagaagg gggaaacc caaaaggcca gcaaaaggcc 2580 ggatacctgt cccctggaag gtgctgcgg tttttccata ggctccgccg cagaggatta cagcactact cagcaggttt cccctggaag tcagcactact ggctccccgcc ccctgagaag ctgcaggggttc cccctcggga agcgtggcgc tttcccaatg gccaccact gccgcttacc 2760 ggatacctgt ccgcctttct ccctcgga agcgtggcgc tttcctaatg gctgttgcg cagagagataccc aggcggtgcgc cacactact ggccactact ggctgttggc gctaccctg gctgttggc ggctgtgcg cacacactact cgccactggc gctttccgc ggtacacacga gctgtgtgca cagagttctt gaagtggtgg ctaacactag gctgtgtgca cagagatctc agagtgcgtg ctacactacg ggcagacacac aaaccaccgc tggtagcggt ggttttttt ttggatacc aaaaccaccgc tggtagcggt ggttttttt ttggaacaa aagagatcta aagagatct taaatcaatt ttaacaaaa atgaagttt aaatcaatct tcacggggt gactcacaca agggctgacacta taaagagttt aaatcaatct tcacggggt gactcacaca agggctgac cagagcgacact tacacagga gaggagggggggggg	cactcatagg	acactcatag	ctcaggaggg	ctccgccttc	aatcccaccc	gctaaagtac	2280
gaagaaatta aagcaagata ggctattaag tgcagaggga gagaaaatgc ctccaacatg 2400 tgaggaggta atgagagaaa tcatagaatt tctccgctt ccctcggtcg ttcggctggt ggagggggta tcagctcact gactcggtg 2520 tccacagaat caggggataa cggaggaaga aacatgtgag caaaaggccg gttsgtggg ttttccata ggctccacaaaa atcgacgct aagtcagagg tggcgaaacc ggcacaaggact ccctcggga gggggggct cccctcggga ggttgtggg tttttccata ggctccactac ggctcccct cccttggga aggagaaccg ggaaaaggcc 2700 caggcgtttc cccctggga ggcggtgggaaacc cccctgggaag cggcggttcc cccttggga aggcggggcg tttccata ggttcagcag ggttgtggg ggtcgttcgc ccctcggga agcggtgggg ctcaactagggttaccaa ggccgtggg cttcagggg ggtgggggggggg	ttggaggggt	ctctccctcc	ctcatcagcc	caccaaacca	aacctagcct	ccaagagtgg	2340
tgaggaagta atgagagaaa tcatagaatt tcttcccact cactegotegotegotegotegotegotegotegotegotego							2400
cgctcggtcg ttcggctgg gcgagcggta tcagctcact caaaggcggt aatacggtta 2580 aggaaccgta aaaaggccg gttgctggcg tttttccata ggctccgcc cactgacgag 2580 catcacaaaa atcgacgct cagtcaggag ttgcgtggcg tttttccata ggctccgcc ccctgacgag 2640 catcacaaaa atcgacgct cagtcaggag tggcgaaacc cgacaggact taaaagatac 2700 aggcgtttc cccctggag ctccctcgtg cgctctctg ttccgaccct gcgaccct ggaaccctg cccctggag ggtgtgtcg cccctggt ggcggcgc tttccaatg ctcacactag ggtacctgt ggtacgttgggaaggtatctca gttcggtgta ggtcgttcgc tccaaggtgg gctgtgtgca ccggaacccac gttaggtga accgtggggggcg tttggtgca ccggaaccactgg gctggtgca accgctggc cttatccggg aactactgg gctgtgtgca ccggaaccactgg gcggtgcta cagagttctt gaagtgggg cctaacactag gctacactag aaggacttat ggcgctctgct gaagccact ggtaacagga ttaggcaaa aaggacagta tttggtact ggcgctctgc ggaagcact ggtaacagga acctcacgggggggcg tggtggaaaaaaa aaggaccagt ggtagcgg ggtttttttg tttggaacaa aaggacagta 3060 cggagaaaaa aaggactca aggagatttt gtgatcttt ctacggggtc tggaagaaaaa accacacgc tggtagcggt ggtttttttt ctacaggggtc gcagattacg 3180 cggaagaaaaa accacacgc tggtagcggt ggtttttttt ctacaggggtc tgacgctcag 3240 cggtcgaaaa accacacgc tggtagcggt ggtgtgttt ctacagagga tagctctaa aaggactga 3360 tggtctgaca gttaccaatac aggactgaat cgcccactaa tccaggaga dcgtctgaa aggagtttt aaatcaatc taaatcaaa atgaaggttt ggaggagggcg tgagtctgc cgtcactacac aggcctgaat cgcccactaa tccagcaga 3480 agggtttggt gatcacaaca ggcctggat tgttgagg ggaggggcgc tgagtctca aagggaga 3540 gccacggttg atgagagact tgttgagaa aaccacacta taaccaatc tgattagaaa aaccacatca ggcctgaat cgccacaca aaccacaca aaccacaca gcgccacacaca	b	auguagaca	backarrak	tattaaaatt	gagaaaacgo	tasatasata	
tccacagaat caggggataa cgcaggaaag aacatgtgag caaaaggcca gcaaaaggcc 2580 aggaaccgta aaaaggccgc gttgctggcg tttttccata ggctccgccc ccctgagag 2640 caccacaaaa atcgacgctc aagtcagagg tggcgaaacc cgacaggact ataaagatac 2700 aggcgtttc ccctcggaa ctccctcgtg cgctctcctg ttccaagcggat ctcaccacacacacacacacacacacacacacacaca	tgaggaagta	atyayayaaa	ccatagaatt	terregeri	Correderrac	tyacttycty	
aggaaccgta aaaaggccgc gttgctggcg tttttccata ggctccgcc Ccctgacgag 2640 catcacaaaa atcgacgctc aagtcagagg tggcgaaacc cgacaggact ataaagatac 2700 caggcgtttc cccctggag ctccctcgtg cgctctcctg ttccgacct gccgttacc 2760 ggatacctg cccctgtgt ccctctgga agcgtgcgc ttttccaatg ctcacgctgt 2820 aggtatctca gttcggcgc cttatccggt accgactgtgt gttcgcccg accgctgcgc cttatccggt accgactat cgcactgcg cttatccggt accgactat cgcactgg accgactgg gttgtgcg ctacactacg gctgtggtac acgagttct gaagtggtgc ctaactacgg gctacactag aaggacgtt tttggtatct gcgctctgc gaagcagct ggatacacgg gctacactag aaggacgta tttggtatct gcgcaaaa aaggactcta aggagtgtgg ggtttttttggtatct gcgctcgc gaagcagt accttcggaa aaaggacgga aacgagaaa aaggactca aggagtttt ggaaggacgaa accacacgg tggtagegg ggtttttttggtatct ttaaattaaa atgaagtttt aaatcaatc aggatttt gtaactatt ttaaattaaa							
agagaccgta aaaaggccgc gttgctggcg tttttccata ggctccgcc ccctgacgag 2640 catcacaaaa atcgacgct aagtcagagg tggcgaaacc cgacaggact ataaagatac 2700 caggcgtttc ccctctgtga ccctctcgtg cgctctcctg ttccgacct gccgcttacc 2760 ggatacctgt ccccctgtga gccgctccctg gccgcttacc 2760 aggtatctca gttcggtgta ggcggtaccc ccctcggga agcggtggcc tttctcaatg ctcacgctgt ggttcagccg accgctgcgc cttatccggt aactatcgtc ttgagtccaa cccaggttagg accggtgcgc cttatccggt aactatcgtc ttgagtccaa cccaggttagg cagagttatc ggcggtgctc cagagtctct gaagtggtgg cctaactacg gctacactag aggggtatgta ggcggtgctccggaaaaa acctaccgc tggagcagtt accttcggaa aaaggaagttgt ttgagcagaa aaaggaatcta aggaagtcct ttgagctcttg ttgagccaaaaaa aggaatctca aggagtttt aaaataaaa atgaagttt aaaacaaatc taaataaaa atgaagttt aaaacaaatc taaataaaa atgaagttt aaaacaaatc taacaatat tacacaata taacaaata taacaaaa gaagaaccta ggagaaatacac ggagaaatacac gaatgaac ggaagaaatacac gaatgaac cagaacaaa aactaacaa gaaaacacaa gaactgaaca gaagaaataaa aactaaaaa gaactaaaaa gaactaaaaa gaactaaaaa gaactaaaaa aactaaaaa gaactaaaaa gaactaaaaa gaactaaaaa aactaacaaa gaacaacacaa aactaaaaa aactaacaaa gaacaacacaa aacaaacaaa aacaaacacaa aacaaacacaa aacaac	tccacagaat	caggggataa	cgcaggaaag	aacatgtgag	caaaaggcca	gcaaaaggcc	2580
catcacaaa atcgacgtt aagtcagagg tggcgaaacc cgacaggact ataaagatac 2700 caggcgtttc cccctggaag ctccctcgtg cgctctcctg ttccgacgct gcgcttacc gcgacatcactg cgccttctc cccttcggaa gcgtggcgc tttctcaatg ctcacacgctgagatactca gttcggtaa ggtcgttcgc bccaagctgg gctgtgtga ccacacgacttat cgccactggc cttatccgg accgactgag gctgtggaagaccacggtgcacacggtgcacacacgg gcaacacacacgactgggtgtacacacacacacgactggataccacacaca	aggaaccgta	aaaaggccgc	attactaaca	tttttccata	gactccaccc	ccctgacgag	2640
cagoguttic cocctogaag ctccotogtg cgcttcctg ttccgacct gccgcttacc ggatacctgt ccgcctttct cccttcggga agcgtggcgc tttctcaatg ctcacgctgt 2820 aggtatctca gttcggtga ggtcgttcgc ccaagctgg gctgtggca cgaaccccc 2880 gttcagccg accgctgcgc cttatccggt aactatcgtc ttgagtccaa cccggtaaga 2940 cacgacttat cgccactggc agagtgggg aactatcgtc ttgagtccaa cccggtaaga 2940 cacgacttat cgccactggc gaggtgggg cactactag ggtagtcgac ggaggtatgta 3000 ggcggtgcta cagagttctt gaaggcagta cctacactag gctacactag gaggatatgta 3120 ttggtatct gcgctctgct gaagccagtt accttcggaa aaaggattgg ttggcccggaaaaa aaccaccgc tggtagcggg ggttttttt ttgcaagca gcagattacg gcagataact ttggaccgaa accacgtta gggatttt ttggtatttt ctacaggggt tgaggcgaa acccactgaa ggagtatgta 3120 ttggacgaaa accacactaa agggatttt ggtcattggaa accacactag gacgattacg ggttttttt ttgaaagaa cccacgtta agggatttt ggtcatgaggg gggggggggg	catcacaaaa	atcoacoctc	aantranann	taacaaacc	caacaagact	ataaagatac	
ggatacctgt ccgccttrct cggtgta ggtcgttcgc ggtcgtcgcg cttatccggt aggtatctca gttcggtgta ggtcgttcgc cttatccggt accgcccg accgctgcg cttatccggt accgactgtg gctgtgtgca cgaaccccc 2880 gttcagcccg accgctgcg cttatccggt accgactgtg gctgtgtgca cgaaccccc 2880 ggtcgttcagccgg accgctgtgtgca agcagcact tttgagtcaa cccggtaaga 2940 ggcggtgcta cagagttctt gaagtggtgg cctaaccaag ggtaacagga ttagcagga gggggtgtatgta 3000 tttggtatct ggcgtctgcg gaggccact gggtaacagga ttagcagga gggtattgta 3000 tccggcaaaca aaaccaccgc tggtagcggt ggttttttt tttgaaaca gaggatctca agaagatcct ttgaatcttt ttgaacatag ggttttttt tttgaaacaa gagatctca aggaagatcct ttgaatcttt ttacaacaaa atgaagttt taaataaaa atgaagttt taaaaaaa ggatctcaa aatgaagttt taaaaaaa ggaggtatgcagg ggggggggcg tgaggtatata tgagtaaact 3360 tggtcagcag gttaccaatg ggggggggcc tgagggtc tgacgctcag 3180 cggtcatacca tagttgcctg accccaatca gagggggggggg							
aggtatctca gttcggtga ggtcgttcgc tccaagctgg gctgttgtca cgaaccccc 2880 gttcagcccg accgctgcgc cttatccggt aactatcgtc ttgagtccaa cccggtaaga 2940 gcggtgtca cagagttctt gaagtgatg ggtaacaaga ttagcagga gaggtatgta 3000 gcggtgtca cagagttctt gaagtgatg cctaactacg gctacactag aaggacagt tttggtatct gcgcactagc tggtagcggt gctacactag aaggacagta 3060 tttggtatct gcgcaccac agaacaccgc tggtagcggt ggttttttg tttgcaagca gcagattacg gcgaaaaaa aaggatctca agaagatcct ttaaattaaa							
gitcagccg accgctgcg citatccggt aactatcgtc ttgagtcaa cccggtaaga gggstatgta 3000 ggcggtgcta cagagtctt gaagtggtgg cctaactacg ttagcagag gaagtatgta 3000 ggcggtgcta cagagtctt gaagtggtgg cctaactacg gtcacactag aaggacagta 3060 tttggtatct gcgctctgct gaagtcggtg cctaactacg gtcacactag aaggacagta 3060 tttggtatct gcgctctgct gaagtcggt ggtttttttg tttgcaagca gcagattacg 3180 cgcagaaaaa aaggactca aggagtttt gtcatgggt ttgacggt ttgacggt tggaagcggt tgaagtcct ttgatcttt ctacggggt ttgacgctcag 3240 tagatccttt taaattaaaa atgaagttt aaatcaatc atacaataa aaggatctt aaggagtttt aaatcaatc atacaagaggtgtgct gaccaatag cttaatcagt gaggggggggc ttgaggtctgc ctcgtgaaga 3480 aggtgttgct gaccaatac aggcctgaat cgccactaa tccaggag aagtgaggag 3540 gccacggttg atgagagctt tgttgaggt ggaccagtt gtgaggtctgc ctcgtgaaga 3480 aggtgttgct atcaacaaa gccgcctgca cgcacata tccagcaaa aagtactat taaccaatt taaccaatt tgattagaaa accacata ttgagagaga accaggtg ggaccagtt tgttgaggag atgcgtaatc ggattttga acttttgctt 3600 tgcacaggaa accaatac caggactat tattcaacaaa gccgccgccc cgtcaagtca gcatcaaatg 3780 taattcaat caggattatc aataccaat ttttgaaaaa gccgtttct taaccaaat caggactat ttttgaaaaa gccgtttctg taatgaaga 3840 gaaaactcac cgaggcagt ccatacaat ttttgaaaaa gccgtttctg taatgaagga 3840 gaaaactcac cgaggcagt ccatacaata caccatataa ttcccctcgt caaaaacg ataccaata accaataaa accaataa cgccgcggaa aagtagagaga 3900 aaaactcac cgaggcagt ccatacaata cgccgcgaga accagtca cacaaaacg 3600 aaaactcac catcaataca accaataaa gccgcttctg taatacaaa 3780 tattcaacaa aaccaataa cacaataaa gccgtttctg taatgaagga 3840 gaaaaactcac cgaggcagt ccatacaaa accaataaa gccgcttccg cagaaacac accaataaa accaataa cacaataaa gccgtttctg taatgagga 3840 gaaaaatcac catcaataa accaataa cggagaagaa aaccaggaa gcaaaaacc catcaaaca accaataaa gccgcaacaa accaacaca 4200 aaaccgtta tcaatcaagg atatcctct aataccaaa aaccaggaa aaccagga							
cacgacttat cgccactggc agcagccact ggtaacagga ttagcagagc gaggtatgta 3000 ggcggtgcta cagagttctt gaagtggtgg cctaactaca ggcacactag aaggacagta 3120 tccggcaaca aaaccaccgc tggtagcggt ggtttttttg tttgcaacaga gaggtatacg 3120 cgcagaaaaa aaggatcca agaagatcct ttgatcttt ctacggggtc tgacgctcag 3180 cgcagaaaaa actcacgtta agaggatttt gtcatgaggt tatcaaaaag gatcttcacc agagtcttt taaattaaaa atgaagttt aaatcaatct taaataata tgagtaaact 3360 tggtcgtgaca gttaccaatg cttaatcaagt ggggggggcgc taaggtctgc ctcgtgaaga 3360 tggtctgaca gttaccaatg cttaatcagt gaggacacta tctcagcgat ctgtcattt 3420 cgtcatcca tagttgcctg actcggggg ggggggggcgc tgaggtctgc ctcgtgaaga 3480 aggtgttgct gactcatacc aggcctgaat cgcccatca tccacgcaa aagtgaggga 3540 ggcacaggtt atgagactt tgttgtaggt ggaccagtt tgaggtctgc ctcgtgaaga 3480 aggtgtggt tgttgaggt ggaccagtt gggggggggg	aggtatctca	gttcggtgta	ggtcgttcgc	tccaagctgg	gctgtgtgca	cgaacccccc	2880
cacgacttat cgccactggc agcagccact ggtaacagga ttagcagagc gaggtatgta 3000 ggcggtgcta cagagttctt gaagtggtgg cctaactaca ggcacactag aaggacagta 3120 tccggcaaca aaaccaccgc tggtagcggt ggtttttttg tttgcaacaga gaggtatacg 3120 cgcagaaaaa aaggatcca agaagatcct ttgatcttt ctacggggtc tgacgctcag 3180 cgcagaaaaa actcacgtta agaggatttt gtcatgaggt tatcaaaaag gatcttcacc agagtcttt taaattaaaa atgaagttt aaatcaatct taaataata tgagtaaact 3360 tggtcgtgaca gttaccaatg cttaatcaagt ggggggggcgc taaggtctgc ctcgtgaaga 3360 tggtctgaca gttaccaatg cttaatcagt gaggacacta tctcagcgat ctgtcattt 3420 cgtcatcca tagttgcctg actcggggg ggggggggcgc tgaggtctgc ctcgtgaaga 3480 aggtgttgct gactcatacc aggcctgaat cgcccatca tccacgcaa aagtgaggga 3540 ggcacaggtt atgagactt tgttgtaggt ggaccagtt tgaggtctgc ctcgtgaaga 3480 aggtgtggt tgttgaggt ggaccagtt gggggggggg	atteagecca	accoctococ	cttatccoot	aactatcotc	ttgagtccaa	cccggtaaga	2940
ggcggtgcta cagagttctt gaagtggtgg cctaactacg gctacactag aaggacagta 3100 tctggcaaac aaaccaccgc tggtagcggt ggtttttttg tttgcaagca gcagattacg 3120 tccggcaaaa aaggatcca agaagacct ttgatcttt ctacggggt tgagctctga 3120 tggaagcaaaa aaccaccgc tggtagcggt ggtttttttg tttgcaagca gcagattacg 3180 tggaacgaaa actcaccgtta aggagatttt gtcatcttt ctacggggtc tgacgctcag 3240 tggaacgaaa actcaccgtta aggagtttt gtcatggaat tatcaaaaaag gatcttcacc 3300 tagatccttt taaattaaaa atgaagttt aaatcaatct aaagtatata tgagtaaact 3360 tggtctgaca gttaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtctattt 3420 cgttcatcca tagttgcctg actccggggg gggggggcgc agggtgttgcc ctcgtgaaga 3480 tgcaccggttg atgagagctt tgttgtaggt ggaaccagttg gtgattttga acttttgctt 3420 tgcaccggaa cggtctgcct tgttgtaggt ggaaccagttg gtgattttga acttttgctt 3600 tgcaccggaa cggtctgcgt tgttgtgagg atgcggtacc tgatcctca actcagcaaa 3660 agttcgatt atcaacaaa gccgccgtcc cgtcaagtca gcgtaatgct ctgccaatg 3720 tacaaccaa taaccaattc tgattagaaa aactcaccaa gcgtacaatg aactcaccaat taaccaattc tgattagaaa aactcaccaa gcgtatctgc tgcgagttc tgattagaaa 3840 gaaaactcac cgaggcagtt ccataggatg gcaaagatcct ggtaacgaa agctgtatccg 3900 actcgtcaa catcaataca acctattaat ttcccctcgt caaaaataaa gctatcaag 3960 aaaccgttat tcaaccaga ccagcaatta cgctgtcaa caaaacaggat ctgaatgcaac aaaccgttat tcatcctga ttggcgctga gcaagaatca cggagaacaac cggagaacaac cggagaacaacac aaaccgtaat tcaaccaga gcagcaataa cgcagcataa aaccgcaaca aaaccgtat tcaatcaga gcagcaataa cgagagaacaacacaac							
thightant gogototgot gaagocagit accitoggaa aaagagitgg tagototiga 3120 tooggaaaaa aaacaacogo tiggtagoggi ggittittig titgcaagoa goagattacg 3180 cigaagaaaa aaccacigaa agaagatot tigatottit ciacggigto tigaagotoo 3240 tigaacigaaa actoacigita agaagatot tigatottit ciacagiggto tigaagotoo 33240 tigaacigaaa actoacigita agaagittig googaaaaaa toogaataaa atgaagitti aaatcaaato tagaagatot tigatotaco 3300 tiggiotigaa gitaccaatig citaatoagi gaggigagoo tigaagoo citaatoo aagitigoo gactoataco aggootigaat cigocoacia toogagoo tigagigoo citaatoo aggootigaat cigocoacia toogagoo aagitigoo atgagigoo gaggigoo tigagigoo citaatoo aggootigaat cigocoacia tigaagoo aagitigoo aagitigoo aggootigaat tigaagago tigagigoo citaatoo tigatoo aggootigaat tigaacaatig gigaatitiga actitigoti 3600 tigacacigga cigocoo cigocoacia aggootigaa aggootigaat tigatoo aaccaatic tigatagaa aaccaatic tigatagaa aaccaatiga googaaaaccaa taaccaatic tigatagaa aaccaatiga googaaaaccaa aaccaatic tigatagaaa aaccaatiga googaaaaccaa aaccaatic tigatagaaa aaccaatiga googaaaaccaa cagagoogagoo caacaaacaa tittigaaaaa googaacaa aaccaatia tittigaaaaa googaacaa aaccaatic tigatagaaa aaccaatic tigaaaaaa googaaaaaccaa cagagoogagoo caacaaacaa tittigaaaaa googaaaacca googaaaaaccaa aaccaatia tittigaaaaa googaaaccaa googaaaaccaa aaccaatia tittigaaaaa googaaaccaa aaccaatia tittigaaaaa googaaaaccaa googaaaaaccaa aaccaatia tittigaaaaa googaaaaaccaa googaaaaaccaa aaccaatia tittigaaaaa googaaaaaccaa googaaaaaccaa googaaaaaccaa googaaaaaccaa googaaaaaccaa googaaaaaccaa googaaaaaccaa aaccaatia tittigaaaaa googaaaaaccaa aaccaatia tittigaaaaa googaaaaaccaa aaccaatia tittigaaaaa googaaaaaccaa caacaacaacaa tittigaaaaa googaaaaccaacaacaacaacaacaacaacaacaacaacaac							
tccggcaaac aaaccaccg tggtageggt ggttttttg tttgcaagca gcagattacg 3180 cgcagaaaaa aaggatctca agaagatcct ttgatcttt ctacggggtc tgacgctcag 3240 tggacgaaa actcacgtta agggattttg gtcatgagat tacaaaaag gatctcacc 3300 tagatccttt taaattaaaa atgaagttt aaatcaatcd aaagtatata tgagtaaact 3360 tggtctgaca gttaccaatg cttaatcagt gaggcaccta tctcaagcgat ctgtctattt cgtcatcacca tagttgcctg actcggggg ggggggggcg tgaggtctgc ctcgtgaaga 3480 aggtgttgct gactcatacc aggcctgaat cgcccatca tccagcaga aagtgaggga 3540 gccacggttg atgagagctt tgttgtaggt ggaccagttg gtgatttga acttttgctt 3600 tgccacggat attcaacaa gccgccgtcc cgtcaagtag gcgtaatgct ctgccaagaa 3660 aggtctgatt attcaacaaa gccgccgtcc cgtcaagtag gcgtaatgct ctgccaagtg 3720 tacaaccaat taaccaatt tgattagaaa aactcatcga gcatcaaatg aaactgcaat 3780 tatattcatat caggattatc aataccatat ttttgaaaaa gccgtttctg taatgaagga 3840 gaaaactcac cgaggcagtt ccataggatg gcaagatcct ggtatcggt tgcgatccg 3900 actcgtccaa catcaataca gaccgaatca ggcagaatga gcaaaagtca acctattaat ttcccctcgt caaaaataag gttatcaagt 3960 aaaccgttat tcattcgtga ttgcgcctga ggcgagacgaa atacggaat gcgaataacc gcgatcaacc atattact tcattcgtga ttgcgcctga gcgagacgaa atacggat gcgaatcacc atacatcac aaccggata aaacaggata aaccggata aaccggata aaccggata aaccggata aaccggata aaccggata aaccggata aaccaggaa atattctca aaccaggaa atattctct aaccaggaa atattctct aaccaggaa atacctgaa ggcagaagaa aacctgcaag ggcagaagaa aaccggata aaccaggaac aaccaggaac aacctgcaag ggcagaaca atacctgaa ggcagaaca atacctgaa ggcagaaca atacctgaa ggcaacaaca ccggcgaaca aacctgccag ggcagaaca aacctgccag ggcagaaca aacctgcaaca aaccaggaac aaccaggaaca aaccagaca aaccagaca aaccagaca aaccaca	ggcggtgcta	cagagilett	gaagtggtgg	cctaactacg	getacactag	aayyacayca	
tggaacgaaa aaggatcta agaagatct ttgatcttt ctacggggtc tgacgctcag 3240 tggaacgaaa actcacgtta agggattttg gtcatgagat tatcaaaaaa gatcttcacc 3300 tagatccttt taaattaaaa atgaagttt aaatcaatct aagtaataa tgagtaaact 3360 tggtctgaca gttaccaatg cttaatcagt gaggaccta tctcagcgat ctgtcattt 3420 cgttcatcca tagttgcctg actccggggg ggggggcgc tgaggtctgc ctcgtgaaga 3480 aggtgttgct gactcatacc aggcctgaat cgcccatca tccagccaga aagtgaggga 3540 gccacggttg atgagggtt tgttgtaggt ggaccagttg gtgattttga acttttgctt 3600 tgccacggaa cggtctgcgt tgttgtgaggt ggaccagttg gtgattttga acttttgctt 3600 tgccacggaa cggtctgcgt tgttgtgaga atgcgtgatc tgatccttca actcagcaaa 3660 tgccacgatt atcaacaaa gccgccgtcc cgtcaagtca ggctaatgct ctgccagtgt tacaaccaat taaccaatat tgattagaaa aactcatcag gcatcaaatg aaactgcaat 3720 tacaaccaat taaccaatat tgattagaaa aactcatcaga gcatcaaatg aaactgcaat 3780 aactcgtccaa catcaataca acctattaat ttcccctcgt caaaaataag gttatcaagga 3900 actcgtccaa catcaataca acctaataat ttcccctcgt caaaaataag gttatcaagt 3960 gagaaatcac catgagtga gaccgatta tcatcagga gggaacgaa atacgggat gcaaaagctt atgcattct ttccagactt gttcaacagg ccagccatta cgctcgtcat caaaatcact cgcatcaacc 4020 aaaccgttat tcattcgtga ttgcgcctga gggaacgaa atacgcgatc gctgttaaaa 4140 ggacaattac aaacaggaat cgaatgcaac cggcggaaga atattcttct aatacctga atgctgtttt cccggggatc gcagtggagggggaataaatt ccgtcagca gttagtcag gacaactca accatcaact atgcatcaaca 4200 atattttcac ctgaatcag atattcttct aatacctgga ataccttgat ggtcgaaga 4320 ctacctttgc catgtttcag aaacactct ggcgcatcaaca atgctgttt cccggggatc atcactctac ctgtaaccac 4200 actattttcac catgagcac gttagtctg accatccac atgggtgaa ataccttcac accatcacac atggggata atattctct aatacctgga ataccttgat ggtcggaaga 4220 accatggtga gacaacatcc ggcagcatca accatccac accatcacac accatcacac accatcacac accatcacacaca	tttggtatct	gcgctctgct	gaagccagtt	accttcggaa	aaagagttgg	tagctcttga	
tggaacgaaa aaggatcta agaagatct ttgatcttt ctacggggtc tgacgctcag 3240 tggaacgaaa actcacgtta agggattttg gtcatgagat tatcaaaaaa gatcttcacc 3300 tagatccttt taaattaaaa atgaagttt aaatcaatct aagtaatata tgagtaaact 3360 tggtctgaca gttaccaatg cttaatcagt gaggaccta tctcagcgat ctgtctattt 3420 cgttcatcca tagttgcctg actccggggg ggggggcgc tgaggtctgc ctcgtgaaga 3480 aggtgttgct gactcatacc aggcctgaat cgcccatca tccagccaga aagtgaggga 3540 gccacggttg atgagggtt tgttgtaggt ggaccagttg gtgattttga acttttgctt 3600 tgccacggaa cggtctgcgt tgttgtgaggt ggaccagttg gtgattttga acttttgctt 3600 tgccacggaa cggtctgcgt tgttgtgaga atgcgtgatc tgatccttca actcagcaaa 3660 tgccacgatt atcaacaaa gccgccgtcc cgtcaagtca ggctaatgct ctgccagtgt tacaaccaat taaccaatat tgattagaaa aactcatcag gcgtaatgct ctgccagtgt 3720 tacaaccaat taaccaatat tgattagaaa aactcatcaga gcgtaatgct tgcgaatga 3780 tattcatat caggattatc aataccatat ttttgaaaaa gccgtttctg taatgaagga 3840 gaaaactcac catgagtgac gactgaatcc ggtgagatct ggtatcaggt 3900 actcgtccaa catcaataca acctattaat ttcccctcgt caaaaataag gttatcaagt 3960 gagaaatcac catgagtgac gactgaatcc ggtgagaatg gcaaaagctt atgcattct ttccagactt gttcaacagg ccagccatta cgctcgtcat caaaatcact cgcatcaaca 4020 aaaccgttat tcattcgtga ttgcgcctga gggagacgaa atacgcgatc gctgttaaaa 4140 ggacaattac aaacaggaat cgaatgcaac cggcggaaga atacctgga atacctgga atacctgga accatctcat ctgtaaccac 4200 atattttcac ctgaatcagg atattcttct aatacctgga ataccttgat ggtcggaaga 4320 ctacctttgc catgtttcag aaacactct ggcgcatca atgggataa atggggata atatcttct cgcatcacc 4200 atattttcac catgagca gttagtctg gcacatcac accatcact attggcaacg 4260 gcagtggtag gacaaattac ccgtcagca gtttagtctg accatccat ctgtaacact attggcaacg 4260 gcagtggtag gacataaatt ccgtcagca gtttagtctg accatccat ctgtaacact attggcaacg 4260 gcagtggtag aaccatgc gtttagtctg accatccat attggcaacg 4260 gcagtggtag accatttcac catggcag gacatcatc acatcactac acatcactac acatcactac acatcactac 4200 attatttcac catgatgcac gtttagtctg accatccat acatcactac acatcactac 4200 attatttcac catgatgcac gtttagtctg accatcactac acatcactac acatcactac 4200 attatttcac catgatgcac ctg	tccggcaaac	aaaccaccgc	tggtagcggt	ggtttttttg	tttgcaagca	gcagattacg	3180
tggaacgaaa actcacgtta agggattttg gtcatgagat tatcaaaaag gatcttcacc tagatccttt taaattaaaa atgaagttt aaatcaatct aaagtatata tgagtaaact 3360 tggtctgaca gttaaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtctattt 3420 cgttcatcca tagttgcctg actccggggg gggggggcgc tagggtctgc ctgtgaaga 3480 gccacggttg atgaggett tgttgtaggt ggaccagttg gtgattttga acttttgctt 3600 tgccacggata atgaggett tgttgtaggt ggaccagttg gtgattttga acttttgctt 3600 tgccacggata cggtctgcgt tgttgtgagg atgcgtgatc tgatccttca actcagcaaa 3660 agttcgatt attcaacaaa gccgccgtcc cgtcaagtca gcgtaatgct ctgccagtgt tacaaccaat taaccaatt tgattagaaa aactcatcga gcatcaaatg aaactgcaat 3780 ttattcatat caggattatc aataccatat ttttgaaaaa gccgtttctg taatgagga 3840 gaaaactcac cgaggcagt ccataggatg gcaagatcct ggtatcggt tgcggattccg actcgtccaa catcaataca acctattaat ttcccctcgt caaaaataag gttatcaagt 3900 gagaaatcac catgagtgac gactgaatcc ggtgagaatg gcaaaagctt atgcattct 4020 ttccagactt tcatcgtga ttgcgcctga gggagacgaa ataccggatc gctgttaaaa 4140 ggacaattac aaacaggaat cgaatgcaac cggcgcagga acactgccag cgcatcaaca 4200 atatttcac ctgaatcagg atattctct aatacctgga atgctttat cacggatag ggacataaatt ccgtcagcag gttagtcag gttagtcaacg gttagtgga gacacatca 4200 atatttcac ccgatcagca gtttagtctg acactctcat ctgaacact atggcgaaga 4320 ggcataaaatt ccgtcagca gtttagtctg acactctcat ctgaacact atggcaacg 4260 ctacctttgc catgtttcag aaacacct ggcgcatcag gcttacacac atcatcaga gtacggataa atgctgttt cccggggatc 4260 ctacctttgc catgtttcag aaacactct ggcgcatcag gcttccata caatcagaa 4380 ctacctttgc catgtttcag aaacactct ggcgcatcgg gcttcccata caatcgaacg 4380 ctacctttgc catgtttcag aaacactct ggcgcatcgg gcttcccata caatcgaacg 4440 attgtcgcac ctgattgccc gacattatcg cgagccatt tatacccata tataccata tataccata 4440 attgtcgcac ctgattgccc gacattatcg cgagccatt tatacccata tataccata 4440 attgtcgcac ctgattgccc gacattatcg cgagccatt tatacccata tataccata 4440							3240
tagatccttt taaattaaaa atgaagtttt aaatcaatct aaagtatata tgagtaaact tggtctgaca gttaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtctattt 3420 cgttcatcca tagttgcctg actccggggg gggggggcc tgaggtctgc ctcgtgaaga 3480 aggtttgct gactcatacc aggcctgaat cgcccatca tccagccaga aagtgaggga 3540 gccacggttg atgagggct tggtgtaggt ggaccagttg gtgattttga acttttgctt tgccacggaa cggtctgcgt tgtgtaggt ggaccagttg gtgattttga acttttgctt tgccacggaa cggtctgcgt tgtgtaggt ggaccagttg gtgattttga acttttgctt actcagcaaa 3660 agttcgatt attcaacaaa gccgccgtcc cgtcaagtca gcgtaatgct ctgccagtgt tgattgagaa aactcatcaa actaataat tgattagaaa aactcatcaa gccgtttctg taatgaagga 3720 tacaaccaat taaccaatt tgattagaaa aactcatcaa gccgtttctg taatgaagga 3840 gaaaactcac caggagtatc cataataat ttttgaaaaa gccgtttctg taatgaagga 3840 gaaaactcac catcaataca acctattaat ttcccctcgt caaaataag gttatcaagt 3960 actcgtccaa catcaataca acctattaat tcccctcgt caaaaataag gttatcaagt 3960 gagaaaatcac catcaacaa acctattaa tccctcgtcat caaaatcact cgcatcaacc 4080 aaaccgttat tcattcgtga ttgcgcctga gcggagaagaa acactgccag cgcatcaaca 4200 atattttcac ctgaatcagg atattcttct aatacctgga atgctgttt cccggggatc gcagtggggggggcc tgagggatcgaa atgcggtta ctgcatcaaca 4200 atattttcac cagaatcag atatcctct aatacctgga accatctcat ctgaacaca 4200 ctacctttgc catgtttcag aaacactct ggcgcatcgg gcttcccata caatcagaa 4380 ctacctttgc catgtttcag aaacactct ggcgcatcgg gcttcccata caatcagaa 4380 ctacctttgc catgtttcag aaacactct ggcgcatcgg gcttcccata caatcagaa 4440 attgtcgcac ctgattgcc gacattatcg cgagccatt tataaccata taaaccata taaaccata 4400 attgtcgcac ctgattgcc gacattatcc caacaccat taaaccata taaaccatca 4400 attgtcgcac ctgattgcc gacattatcc catggcata accatccata taaaccata 4400 attgtcgcac ctgattgcc catgattcc ctgaacact 4400 attgtcgcac ctgattgcc catgatcc ctgattgccat taaaccata 4400 attgtcgcac ctgattgcc ctgattgcc catgatccat taaaccata 4400 attgtcgcac ctgattgccc catgatcc ctgatccata 4400 attgtcgcac ctgattgccc catgatcc ctgatccata 4400 attgtcgcac ctgattgccc catgatccc ctgatccata 4400 attgtcgcac ctgattgccc ctgatcacat 4400 attgtcgcac ctgattgccc catgatccc ctgatcca							
tggtctgaca gttaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtctattt cgttcatcca tagttgcctg actccggggg gggggggcc tgaggtctgc ctcgtgaaga 3480 aggtgttgct gactcatacc aggcctgaat cgcccatca tccagccaga aagtgaggga 3540 gccacggttg atgaggact tgttgtaggt ggaccagttg gtgattttga acttttgctt 3600 tgccacggaa cggtctgcgt tgtcgggaag atgcgtgatc tgatccttca actcagcaaa 3660 agttcgatt attcaacaaa gccgccgtcc cgtcaagtca gcgtaatgct ctgccagtgt tgatcggaag aactcatcaga gcgtaatgct ctgccagtgt 3720 tacaaccaat taaccaatt tgattagaaa aactcatcga gcgtaatgct ctgccagtgt 3720 tacaaccaat caggattatc aataccatat ttttgaaaaa gccgtttctg taatgaagga 3840 gaaaactcac cgaggcagtt ccataggatg gcaagatcct ggtatcggtc tgcgattccg 3900 actcgtccaa catcaataca acctattaat ttccatcat ggtaatcggt tgcgaatcgg gacaaactcac catgagtgac gactgaatcc ggtgagaatg gcaaaagctt atgcatttct 4020 ttccagactt gttcaacagg ccagccatta cgctcgtcat caaaaatcact cgcatcaaca 4080 aaaccgttat tcattcgtga ttgcgcctga gcgagacgaa atacctgga atacctgga gtaaccatga gtaacaacac actacaagg gtaaccagg atattcttct aatacctgga atgctgatc gcgatcaaca 4200 atattttcac ctgaatcag gtaaccatca gttaggtaa accattcat cattggaaga ggcataaatt ccgtcagca gttagtctg accatctcat ctgaacaca atcactttgc catgtttcag aaacactc ggcgcatca ctgaacac ctgaacaca 4200 ctacctttgc catgtttcag aaacactc ggcgcatcat catcacac acatccatc attggcaacg 4380 ctacctttgc catgtttcag aaacactc ggcgcatcat tatacccata tataccata attggcaacg 4440 attgtcgcac ctgattgccc gacattatcg cgagccatt tatacccata tataccata tataccata 4440 attgtcgcac ctgattgccc gacattatcg cgagccatt tatacccata tataccata 4400 attgtcgcac ctgattgccc gacattatcg cgagccatt tatacccata tataccata 4440 attgtcgcac ctgattgccc gacattatcg cgagccatt tatacccata tatacccata 4400 attgtcgcac ctgattgccc gacattatcg cgagcccatt tatacccata tatacccata 4400 attgtcgccc attgtcgcc gacattatcg cgagccc							
cgttcatcca tagttgcctg actccggggg gggggggcgc tgaggtctgc ctcgtgaaga 3540 gcacggttg gactcatacc aggcctgaat cgcccatca tccagccaga aagtgaggga 3540 gcacggttg atgagagett tgttgtaggt ggaccagttg gtgattttga acttttgctt 3600 tgccacggaa cggtctgcgt tgtcgggaag atgcgtgatc tgatccttca actcagcaaa 3660 agttcgattt attcaacaaa gccgccgtcc cgtcaagtca gcgtaatgct ctgccagtgt tattcaacaat taaccaatt tgattagaaa aactcatcag gcatcaaatg aaactgcaat 3720 tacaaccaat taaccaatt tgattagaaa aactcatcag gcatcaaatg aaactgcaat 3780 gaaaactcac cgaggcagtt ccataggatg gcaagatcct ggtatcggt tgcggattccg 3900 actcgtccaa catcaataca acctattaat ttcccctcgt caaaaataag gttatcaagt 3960 gagaaatcac catgagtgac gactgaatcc ggtgagaatg gcaaaagctt atgcattct 4020 ttccagactt gttcaacagg ccagccatta cgctcgtcat caaaaatcact cgcatcaacc 4080 aaaccgttat tcattcgtga ttgcgcctga gcgagaagaa atacgcgatc gcgttaaaa 4140 ggacaattac aaacaggaat cgaatgcaac cggcgcagga acactgccag cgcatcaaca 4200 gcatgagtgg gtaaccatgc atcatcagga gtaacggaaa atactctcat ctgtaacatc ccggggatc gcatcaact 4260 gcagtggtga gtaaccatgc atcatcagga gcatcactca ctgtaacatc cgctcgaaca 4220 gcataaaatt ccgtcagcca gtttagtctg acactctcat ctgtaacatc attggcaacg 4320 gcataaaatt ccgtcagcca gtttagtctg acactctcat ctgtaacatc attggcaacg 4320 ctacctttgc catgtttcag aaacaactct ggcgcatcag gcttcccata caatcgaag 4320 ctacctttgc catgtttcag aaacaactct ggcgcatcatt tatacccata taaatcagca 4440 attgtcgcac ctgattgccc gacattatcg cgagccatt tatacccata taaatcagca 4500							
aggtgttgct gactcatacc aggcctgaat cgcccatca tccagccaga aagtgaggga 3540 gccacggttg atgagagct tgttgtaggt ggaccagttg gtgattttga acttttgctt 3600 tgccacggaa cggtctgcgt tgtcgggaag atgcgtgatc tgatccttca actcagcaaa 3660 agttcgatt tatcaacaaa gccgccgtcc cgtcaagtca gcgtaatgct ctgccagtgt 3720 tacaaccaat taaccaattc tgattagaaa aactcatcga gcatcaaatg aaactgcaat 3780 gaaaactcac cgaggcagtt ccataggatg gcaaaaatg gcatcaaatg 3840 gaaaactcac cgaggcagtt ccataggatg gcaaaattact ggtatcggtc tgcgattccg 3900 actcgtccaa catcaataca acctattaat ttcccctcgt caaaaataag gttatcaagt 3960 gagaaatcac catgagtgac gactgaatcc ggtgagaatg gcaaaagctt atgcattct 4020 ttccagactt gttcaacagg ccagccatta cgctcgtcat caaaatcact cgcatcaacc 4080 aaaccgttat tcattcgtga ttgcgcctga gcgagacgaa atacgcgatc gcgatcaaca 4200 atatttcac ctgaatcagg atattctct aatacctgga acactgccag cgcatcaaca 4200 gcagtggtga gtaaccatgc atcatcagg gtacggataa atgctttat ccggggaga 4320 gcataaaatt ccgtcagcca gtttagtctg acactctcat ctgtaacatc aatcgatag 4380 ctacctttgc catgtttcag aaacaactc ggcgcatcgg gcttcccata caatcgatag 4440 attgtcgcac ctgattgccc gacattatcg cgagccatt tatacccata taaatcagca 4500	tggtctgaca	gttaccaatg	cttaatcagt	gaggcaccta	tctcagcgat	ctgtctattt	3420
aggtgttgct gactcatacc aggcctgaat cgcccatca tccagccaga aagtgaggga 3540 gccacggttg atgagagctt tgttgtaggt ggaccagttg gtgattttga acttttgctt 3600 tgccacggaa cggtctgcgt tgtcgggaag atgcgtgatc tgatccttca actcagcaaa 3660 agttcgattt atcaacaaa gccgccgtcc cgtcaagtca gcgtaatgct ctgccagtgt tacaaccaat taaccaattc tgattagaaa aactcatcag gcatcaaatg aaactgcaat 3720 tacaaccaat taaccaattc tgattagaaa aactcatcag gcatcaaatg aaactgcaat 3780 gaaaactcac cgaggcagtt ccataggatg gcaagatcct ggtatcggtc tgcgattccg 3900 actcgtccaa catcaataca acctattaat ttcccctcgt caaaaataag gttatcaagt 3960 gagaaatcac catgagtgac gactgaatcc ggtgagaatg gcaaaagctt atgcattct 4020 ttccagactt ggtcaacagg ccagccatta cgctcgtcat caaaaatcact cgcatcaacc 4080 aaaccgttat tcattcgtga ttgcgcctga gcgagaagaa atacgcgatc gcgatcaaca 4200 gcatgagtga gtaaccatgc atcatcagga gtaaccatga ggcataaatt ccggaggaa atactctcat ctgtaacaca aatgctgata ggacataaatt ccgcagcag atattctct acaacctggaa ggacataaatt ccgcagcag gttaggaaa aatgcttgat ggtcggaaga 4320 ggcataaaatt ccgtcagcca gtttagtctg acaactcatc ctgtaacata attgcaacg 4440 attgtcgcac ctgattgcc gacattatcg cgagccatt tatacccata taaatcagca 4500	cattcatcca	tagttgcctg	actccggggg	aaaaaaacac	tgaggtctgc	ctcqtqaaqa	3480
gccacggttg atgagagett tgttgtaggt ggaccagttg gtgattttga acttttgett 3600 tgccacggaa cggtctgcgt tgtcgggaag atgcgtgate tgatcettca actcageaaa 3660 agttcgattt attcaacaaa gccgccgtce cgtcaagtca gcgtaatget ctgccagtgt 3720 tacaaccaat taaccaatte tgattagaaa aactcatega gcatcaaatg aaactgcaat 3780 tattcatat caggattate aataccaat ttttgaaaaa gccgtttetg taatgaagga 3840 gaaaactcae cgaggcagte ccataggatg gcaagateet ggtateggte tgcgattceg 3900 actcgtccaa catcaataca acctattaat ttcccctcgt caaaaataaag gttatcaagt 3960 gagaaatcae catgagtgae gactgaatee ggtgagaatg gcaaaagett atgcatteet 4020 tccagactt gttcaacagg ccagccatta cgctcgtcat caaaaatcact cgcatcaace 4080 aaaccgttat tcattcgtga ttgcgcctga gcgagacgaa atacgcgate gctgttaaaa 4140 ggacaattac aaacaggaat cgaatgcaac cggcgcagga acactgccag cgcatcaaca 4200 atatttcac ctgaatcagg atattetet aatacctgga atgctgtte cccggggate 4320 ggcataaaatt ccgtcagca gtttagtctg acactctcat ctgtaacate attgcaacg 4380 ctacctttge catgtttcag aaacaactct ggcgcatcgg gcttcccata caatcgatag 4440 attgtcgcac ctgattgccc gacattatcg cgagccatt tatacccata taaatcagca 4500							3540
tgccacggaa cggtctgcgt tgtcgggaag atgcgtgatc tgatccttca actcagcaaa agttcgattt attcaacaaa gccgccgtcc cgtcaagtca gcgtaatgct ctgccagtgt 3720 tacaaccaat taaccaattc tgattagaaa aactcatcga gcatcaaatg aaactgcaat 3780 ttattcataat caggattatc aataccatat ttttgaaaaaa gccgtttctg taatgaagga 3840 gaaaactcac cgaggcagtt ccataggatg gcaagatcct ggtatcggtc tgcgattccg actcgtccaa catcaataca acctattaat ttcccctcgt caaaaataag gttatcaagt 3900 gagaaatcac catgagtgac gactgaatcc ggtgagaatg gcaaaagctt atgcattct 4020 ttccagactt gttcaacagg ccagccatta cgctcgtcat caaaatcact cgcatcaacc 4080 aaaccgttat tcattcgtga ttgcgcctga gcgagaagaa atacgcgatc gctgttaaaa 4140 ggacaattac aaacaggaat cgaatgcaac cggcgcagga acactgccag cgcatcaaca 4200 atattttcac ctgaatcagg atattcttct aatacctgga atgctgtttt cccggggatc gcagtggtga gtaaccatgc atcatcagga gtaaggata acactgctag ggtcgaagaa 4320 ggacataaatt ccgtcagca gtttagtctg accatctcat ctgtaacatc attggcaacg 4380 ctacctttgc catgtttcag aaacaactct ggcgcatcgg gcttcccata caatcgatag 4440 attgtcgcac ctgattgccc gacattatcg cgagccatt tatacccata taaatcagca							
agttcgattt attcaacaaa gccgccgtcc cgtcaagtca gcgtaatgct ctgccagtgt tacaaccaat taaccaattc tgattagaaa aactcatcga gcatcaaatg aaactgcaat 3780 ttattcatat caggattatc aataccatat ttttgaaaaa gccgtttctg taatgaagga 3840 gaaaactcac cgaggcagtt ccataggatg gcaagatct ggtatcggtc tgcgattccg 3900 actcgtccaa catcaataca acctattaat ttcccctcgt caaaaataag gttatcaagt 3960 gagaaatcac catgagtgac gactgaatcc ggtgagaatg gcaaaagctt atgcatttct 4020 ttccagactt gttcaacagg ccagccatta cgctcgtcat caaaatcact cgcatcaacc 4080 aaaccgttat tcattcgtga ttgcgcctga gcgagacgaa atacgcgatc gctgttaaaa 4140 ggacaattac aaacaggaat cgaatgcaac cggcgcagga acactgccag cgcatcaaca 4200 atattttcac ctgaatcagg atattcttct aatacctgga atgctgtttt cccggggatc gcaggaggatc gcagtggtga gtaaccatgc atcatcagg gtaagataaatt ccgtcagca gttagtctg accatctcat ctgtaacatc attggcaacg 4380 ctacctttgc catgtttcag aaacaactct ggcgcatcgg gcttcccata caatcgatag 4440 attgtcgcac ctgattgccc gacattatcg cgagcccatt tatacccata taaatcagca 4500							
tacaaccaat taaccaattc tgattagaaa aactcatcga gcatcaaatg aaactgcaat 3780 ttattcatat caggattatc aataccatat ttttgaaaaa gccgtttctg taatgaagga 3840 gaaaactcac cgaggcagtt ccataggatg gcaagatcct ggtatcggtc tgcgattcccg actcgtccaa catcaataca acctattaat ttcccctcgt caaaaataag gttatcaagt 3960 gagaaatcac catgagtgac gactgaatcc ggtgagaatg gcaaaagctt atgcatttct 4020 ttccagactt gttcaacagg ccagccatta cgctcgtcat caaaatcact cgcatcaacc 4080 aaaccgttat tcattcgtga ttgcgcctga gcgagacgaa atacgcgatc gctgttaaaa 4140 ggacaattac aaacaggaat cgaatgcaac cggcgcagga acactgccag cgcatcaaca 4200 atattttcac ctgaatcagg atattcttct aatacctgga atgctgtttt cccggggatc gcagtggtgg gtaaccatgc atcatcagga gtaaggata aatgctgtttt cccggggatc gcgatgatggggataaatt ccgtcagca gtttagtctg accatctcat ctgaacact attggcaacg 4380 ctacctttgc catgtttcag aaacaactct ggcgcatcg gcttcccata caatcgatag 4440 attgtcgcac ctgattgccc gacattatcg cgagcccatt tatacccata taaatcagca							
tacaaccaat taaccaattc tgattagaaa aactcatcga gcatcaaatg aaactgcaat 3780 ttattcatat caggattatc aataccatat ttttgaaaaa gccgtttctg taatgaagga 3840 gaaaactcac cgaggcagtt ccataggatg gcaagatcct ggtatcggtc tgcgattcccg actcgtccaa catcaataca acctattaat ttcccctcgt caaaaataag gttatcaagt 3960 gagaaatcac catgagtgac gactgaatcc ggtgagaatg gcaaaagctt atgcatttct 4020 ttccagactt gttcaacagg ccagccatta cgctcgtcat caaaatcact cgcatcaacc 4080 aaaccgttat tcattcgtga ttgcgcctga gcgagacgaa atacgcgatc gctgttaaaa 4140 ggacaattac aaacaggaat cgaatgcaac cggcgcagga acactgccag cgcatcaaca 4200 atattttcac ctgaatcagg atattcttct aatacctgga atgctgtttt cccggggatc gcagtggtgg gtaaccatgc atcatcagga gtaaggata aatgctgtttt cccggggatc gcgatgatggggataaatt ccgtcagca gtttagtctg accatctcat ctgaacact attggcaacg 4380 ctacctttgc catgtttcag aaacaactct ggcgcatcg gcttcccata caatcgatag 4440 attgtcgcac ctgattgccc gacattatcg cgagcccatt tatacccata taaatcagca	agttcgattt	attcaacaaa	gccgccgtcc	cgtcaagtca	gcgtaatgct	ctgccagtgt	3720
ttattcatat caggattatc aataccatat ttttgaaaaa gccgtttctg taatgaagga 3840 gaaaactcac cgaggcagtt ccataggatg gcaagatcct ggtatcggtc tgcgattccg 3900 actcgtccaa catcaataca acctattaat ttcccctcgt caaaaataag gttatcaagt 3960 gagaaatcac catgagtgac gactgaatcc ggtgagaatg gcaaaagctt atgcatttct 4020 ttccagactt gttcaacagg ccagccatta cgctcgtcat caaaaatcact cgcatcaacc 4080 aaaccgttat tcattcgtga ttgcgcctga gcgagacgaa atacgcgatc gctgttaaaa 4140 ggacaattac aaacaggaat cgaatgcaac cggcgcagga atgctgccag cgcatcaaca 4200 atattttcac ctgaatcagg atattctct atacctgga atgctgttt cccggggatc 4260 gcagtggtga gtaaccatgc attatgctg accatctcat ctgtaacatc attggcaacg 4380 ctacctttgc catgtttcag aaacaactct ggcgcatcag gcttcccata caatcgatag 4440 attgtcgcac ctgattgccc gacattatcg cgagcccatt tatacccata taaatcagca 4500							3780
gaaaactcac cgaggcagtt ccataggatg gcaagatcct ggtatcggtc tgcgattccg 3900 actogtccaa catcaataca acctattaat ttcccctcgt caaaaataag gttatcaagt 3960 gagaaatcac catgagtgac gactgaatcc ggtgagaatg gcaaaagctt atgcattct 4020 ttccagactt gttcaacagg ccagcatta cgctcgtcat caaaaatcact cgcatcaacc 4080 aaaccgttat tcattcgtga ttgcgcctga gcgagacgaa atacgcgatc gctgttaaaa ggacaattac aaacaggaat cgaatgcaac cggcgcagga acactgccag cgcatcaaca 4200 atattttcac ctgaatcagg atattctct aatacctgga atgctgttt cccggggatc gcgatgaaga cgagtggtga gtaaccatga atcatcagga gtaacggataa aatgcttgat ggtcggaaga 4320 ggcataaaatt ccgtcagcca gtttagtctg accatctcat ctgtaacatc attggcaacg 4380 ctacctttgc catgtttcag aaacaactct ggcgcatcat tatacccata taaatcagca 4500							
actcgtccaa catcaataca acctattaat ttcccctcgt caaaaataag gttatcaagt gagaaatcac catgagtgac gactgaatcc ggtgagaatg gcaaaagctt atgcattct 4020 ttccagactt gttcaacagg ccagccatta cgctcgtcat caaaatcact cgcatcaacc 4080 aaaccgttat tcattcgtga ttgcgcctga gcgagacgaa atacgcgatc gctgttaaaa 4140 ggacaattac aaacaggaat cgaatgcaac cggcgcagga acactgccag cgcatcaaca 4200 atattttcac ctgaatcagg atattctct aatacetgga atgctgttt cccggggatc gcagtggtga gtaaccatgc atcatcagga gtaacggataa aatgcttgat ggtcggaaga 4320 ggcataaaatt ccgtcagcca gtttagtctg accatctcat ctgtaacatc attggcaacg 4380 ctacctttgc catgtttcag aaacaactct ggcgcatcgg gcttcccata caatcgatag 4440 attgtcgcac ctgattgccc gacattatcg cgagcccatt tatacccata taaatcagca							
gagaaatcac catgagtgac gactgaatcc ggtgagaatg gcaaaagctt atgcattct 4020 ttccagactt gttcaacagg ccagccatta cgctcgtcat caaaatcact cgcatcaacc 4080 aaaccgttat tcattcgtga ttgcgcctga gcgagacgaa atacgcgatc gctgttaaaa 4140 ggacaattac aaacaggaat cgaatgcaac cggcgcagga acactgccag cgcatcaaca 4200 atatttcac ctgaatcagg atattctct aatacctgga atgctgttt ccggggatc 4260 gcagtggtga gtaaccatgc atcatcagga gtacggataa aatgcttgat ggtcggaaga 4320 ggcataaaatt ccgtcagca gtttagtctg acatctcat ctgtaacatc attggcaacg 4380 ctacctttgc catgtttcag aaacaactct ggcgcatcgg gcttcccata caatcgatag 4440 attgtcgcac ctgattgccc gacattatcg cgagcccatt tatacccata taaatcagca 4500							
ttccagactt gttcaacagg ccagccatta cgctcgtcat caaaatcact cgcatcaacc 4080 aaaccgttat tcattcgtga ttgcgcctga gcgagacgaa atacgcgatc gctgttaaaa 4140 ggacaattac aaacaggaat cgaatgcaac cggcgcagga acactgccag cgcatcaaca 4200 atattttcac ctgaatcagg atattcttct aatacctgga atgctgttt cccggggatc gcagtggtga gtaaccatgc atcatcagga gtacggataa aatgcttgat ggtcggaaga 4320 ctacctttgc catgttcag aaacactct ggcgcatcgg gcttcccata caatcgaag 4380 ctacctttgc catgttcag aaacactct ggcgcatcgg gcttcccata caatcgaag 4440 attgtcgcac ctgattgccc gacattatcg cgagcccatt tatacccata taaatcagca	actcgtccaa	catcaataca	acctattaat	ttcccctcgt	caaaaataag	gttatcaagt	3960
ttccagactt gttcaacagg ccagccatta cgctcgtcat caaaatcact cgcatcaacc 4080 aaaccgttat tcattcgtga ttgcgcctga gcgagacgaa atacgcgatc gctgttaaaa 4140 ggacaattac aaacaggaat cgaatgcaac cggcgcagga acactgccag cgcatcaaca 4200 atattttcac ctgaatcagg atattcttc aatacctgga atgctgttt cccggggatc 4260 gcagtggtga gtaaccatgc atcatcagga gtacggataa aatgcttgat ggtcggaaga 4320 ggcataaaatt ccgtcagca gtttagtctg accatctcat ctgtaacatc attggcaacg 4380 ctacctttgc catgtttcag aaacaactct ggcgcatcgg gcttcccata caatcgatag 4440 attgtcgcac ctgattgccc gacattatcg cgagcccatt tatacccata taaatcagca 4500	gagaaatcac	catgagtgac	gactgaatcc	ggtgagaatg	gcaaaagctt	atgcatttct	4020
aaaccgttat tcattcgtga ttgcgcctga gcgagacgaa atacgcgatc gctgttaaaa ggacaattac aaacaggaat cgaatgcaac cggcgcagga acactgccag cgcatcaaca 4200 atattttcac ctgaatcagg atattcttct aatacctgga atgctgttt cccggggatc gcagtggtga gtaaccatgc atcatcagga gtacggataa aatgcttgat ggtcggaaga 4320 ggcataaatt ccgtcagcca gtttagtctg accatctcat ctgtaacatc attggcaacg 4380 ctacctttgc catgtttcag aaacaactct ggcgcatcagg gcttcccata caatcgatag 4440 attgtcgcac ctgattgccc gacattatcg cgagcccatt tatacccata taaatcagca 4500							4080
ggacaattac aaacaggaat cgaatgcaac cggcgcagga acactgccag cgcatcaaca 4200 atattttcac ctgaatcagg atattcttct aatacctgga atgctgttt cccggggatc 4260 gcagtggtga gtaaccatgc atcatcagga gtacggataa aatgcttgat ggtcggaaga 4320 ggcataaatt ccgtcagcca gtttagtctg accatctcat ctgtaacatc attggcaacg 4380 ctacctttgc catgttcag aaacaactct ggcgcatcagg gcttcccata caatcgatag 4440 attgtcgcac ctgattgccc gacattatcg cgagcccatt tatacccata taaatcagca 4500							
atattttcac ctgaatcagg atattcttct aatacctgga atgctgttt cccggggatc gcagtggtga gtaaccatgc atcatcagga gtacggataa aatgcttgat ggtcggaaga 4320 ggcataaatt ccgtcagcca gtttagtctg accatctcat ctgtaacatc attggcaacg 4380 ctacctttgc catgtttcag aaacaactct ggcgcatcgg gcttcccata caatcgatag 4440 attgtcgcac ctgattgccc gacattatcg cgagccatt tatacccata taaatcagca 4500	adaccyctat	cacciguga	cogogootya	gcgagacgaa	acacycyacc	gotyttaaad	
gcagtggtga gtaaccatgc atcatcagga gtacggataa aatgcttgat ggtcggaaga 4320 ggcataaatt ccgtcagcca gtttagtctg accatctcat ctgtaacatc attggcaacg 4380 ctacctttgc catgtttcag aaacaactct ggcgcatcgg gcttcccata caatcgatag 4440 attgtcgcac ctgattgccc gacattatcg cgagcccatt tatacccata taaatcagca 4500							-
gcagtggtga gtaaccatgc atcatcagga gtacggataa aatgcttgat ggtcggaaga 4320 ggcataaatt ccgtcagcca gtttagtctg accatctcat ctgtaacatc attggcaacg 4380 ctacctttgc catgtttcag aaacaactct ggcgcatcgg gcttcccata caatcgatag 4440 attgtcgcac ctgattgccc gacattatcg cgagcccatt tatacccata taaatcagca 4500							
ggcataaatt ccgtcagcca gtttagtctg accatctcat ctgtaacatc attggcaacg 4380 ctacctttgc catgtttcag aaacaactct ggcgcatcgg gcttcccata caatcgatag 4440 attgtcgcac ctgattgccc gacattatcg cgagcccatt tatacccata taaatcagca 4500							4320
ctacctttgc catgtttcag aaacaactct ggcgcatcgg gcttcccata caatcgatag 4440 attgtcgcac ctgattgccc gacattatcg cgagcccatt tatacccata taaatcagca 4500							
attgtcgcac ctgattgccc gacattatcg cgagcccatt tatacccata taaatcagca 4500							
tccatgttgg aatttaatcg cggcctcgag caagacgttt cccgttgaat atggctcata 4560							
	tccatgttgg	aatttaatcg	cggcctcgag	caagacgttt	cccgttgaat	atggctcata	4560

```
acaccccttg tattactgtt tatgtaagca gacagtttta ttgttcatga tgatatattt
                                                                               4680 ·
ttatcttgtg caatgtaaca tcagagattt tgagacacaa cgtggctttc ccccccccc
cattattgaa gcatttatca gggttattgt ctcatgagcg gatacatatt tgaatgtatt tagaaaaata aacaaatagg ggttccgcgc acatttcccc gaaaagtgcc acctgacgtc
                                                                               4740
                                                                               4800
taagaaacca ttattatcat gacattaacc tataaaaata ggcgtatcac gaggcccttt
                                                                               4860
                                                                               4864
<210> 26
<211> 36
<212> DNA
<213> Artificial Sequence
<220>
<223> oligonucleotide
<400> 26
                                                                                 36
ggtacaagat ctccgccccc atctccccca ttgaga
<210> 27
<211> 33
<212> DNA
<213> Artificial Sequence
<220>
<223> oligonucleotide
<400> 27
ccacatagat ctgcccgggc tttagtcctc atc
                                                                                 33
<210> 28
<211> 27
<212> PRT
<213> Homo sapien
<400> 28
Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly
1. 10 15
Ala Val Phe Val Ser Pro Ser Glu Ile Ser Ser
<210> 29
<211> 45
<212> DNA
<213> Artificial Sequence
<220>
<223> oligonucleotide
<400> 29
caggegagat ctaccatggc ccccattagc cctattgaga ctgta
                                                                                 45
<210> 30
<211> 48
<212> DNA
<213> Artificial Sequence
<220>
<223> oligonucleotide
caggegagat etgeeeggge tttaateete ateetgteta ettgeeac
                                                                                 48
```

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US00/34724

A. CLASSIFICATION OF SUBJECT MATTER  IPC(7): A61K 48/00; C12Q 1/70.  US CL: 514/44; 435/5; 424/93.1.  According to International Patent Classification (IPC) or to both national classification and IPC  B. FIELDS SEARCHED  Minimum documentation searched (classification system followed by classification symbols)  U.S.: 514/44; 435/5; 424/93.1.							
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched							
Electronic data base consulted during the international search (n Medline, embase, scisearch, biosis, caplus and WEST	ame of data base and, where practicable, s	earch terms used)					
C. DOCUMENTS CONSIDERED TO BE RELEVANT							
Category * Citation of document, with indication, where		Relevant to claim No.					
Y,P US 6,099,848 A (FRANKEL et al) 08 August 200	0 (08.08.2000), page 12 paragraph 6.	1-14, 17					
Y WO 97/31115 A2 (MERCK & CO. INC.), 28 Au	gust 1997, page 36.	4					
X WO 90/10230 A1 (UNIVERSITY OF OTTAWA)	07 September 1990, page 11.	17					
Y US 5,858,646 A (KANG) 12 January 1999 (12.01	• '. '	1-14, 17					
	.1555), con. 2, mics 10-17	1-14, 17					
· ·							
		·					
·	•						
	, ,						
	·						
Further documents are listed in the continuation of Box C.	See patent family annex.						
Special categories of cited documents:	"T" later document published after the inter						
"A" document defining the general state of the art which is not considered to be of particular relevance	date and not in conflict with the application of theory underlying the investigation of the conflict with the application of the conflict with the conflict with the application of the conflict with the conflict with the conflict with the application of the conflict with the co						
"E" earlier application or patent published on or after the international filing date	"X" document of particular relevance; the considered novel or cannot be consider						
"L" document which may throw doubts on priority claim(s) or which is clied to establish the publication date of another citation or other special reason (as specified)	when the document is taken alone  "Y"  document of particular relevance; the o						
"O" document referring to an oral disclosure, use, exhibition or other means	considered to involve an inventive step combined with one or more other such being obvious to a person skilled in the	documents, such combination					
"P" document published prior to the international filing date but later than the priority date claimed	"&" document member of the same patent f	amity					
Date of the actual completion of the international search	Date of mailing of the international sear	ch report					
22 February 2001	09 WAR 200 <u>1</u>	on report					
Name and mailing address of the ISA/US	Authorized officer	VI DEV ACT					
Commissioner of Patents and Trademarks	1610	Y J. DEY					
Box PCT Washington, D.C. 20231	Eleanor Sorbello	CENTER 1600					
Facsimile No. (703)305-3230	Telephone No. 703-308-0196	OCIVIED 1000					

Form PCT/ISA/210 (second sheet) (July 1998)

## INTERNATIONAL SEARCH REPORT

Internat application No.

PCT/US00/34724

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)					
This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:					
1. Claim Nos.: because they relate to subject matter not required to be searched by this Authority, namely:					
Claim Nos.:  because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:					
3. Claim Nos.: 15 & 16 because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).					
Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)					
This International Searching Authority found multiple inventions in this international application, as follows:					
1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.					
2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.					
As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:					
4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:					
Remark on Protest The additional search fees were accompanied by the applicant's protest.					
No protest accompanied the payment of additional search fees.					
·					

Form PCT/ISA/210 (continuation of first sheet(1)) (July 1998)